Artifact Intelligence: A Proposed Method to Develop a New Classification System

Brayden Kohler

Abstract

The artifact classification systems that modern archeologists use can take a long time for a human to create and analyze. Moreover, they are not computationally grounded, being based on modern perspectives of the finds. The right classification system is important because it conveys useful information. Many times, people care more about a vessel's appearance than its origins. A new typology system may reveal this relationship. A way to efficiently generate and analyze new systems using mathematical methods are on the horizon of this field. In this paper I use a dataset of Chinese Neolithic pottery images sorted into the categories, Banpo, Banshan, Machang, Majiayao, and Miaodigou, their provenances. I use a convolutional neural network trained on ImageNet to transform these images to vectors, and reduce this data to ten dimensions using a principal component analysis. These vectors are then clustered using k-means and different clusterings are scored using the silhouette score and Davies Bouldin index. These metrics reveal that the images can be clustered into four categories. This new categorization increases the silhouette score of the clustering by 311% than when clustered based on provenance. Further analysis on these clusters reveal relationships between Chinese Neolithic sites: ~90% of the pottery from Miaodigou is placed in one cluster and so is < 5% of Banshan pottery, however, <5% of Miaodigou is in another cluster and nearly 40% of Banshan pottery is too. Archeologists, therefore, can use this method to quickly develop new artifact typologies and reveal or justify patterns between archeological sites.

Key Words

Ceramics, Clustering, CNN, Typology,

Introduction

Traditional artifact classification systems in archaeology, while foundational to the field, are often time-intensive to create and analyze, and they rely heavily on subjective human perspectives. These systems can be proven to be very helpful and are widely used. Pawlowicz and Downum (2021) have shown the effectiveness of using convolutional neural networks (CNN) in classifying pottery into predefined categories. Further, they suggested that their model can show that a suggested type might not be a meaningful type to have in a given system. CNNs (LeCun 1998) are neural networks that identify features in a dataset by repeatedly filtering the data to extract meaningful parts of it. A popular use, then, of CNNs is in image recognition. Chetouani et al. (2020) have also shown the effectiveness of using CNNs in classifying pottery based on the designs engraved upon them. Horn et. al. (2022) showed that CNNs can also be used to identify and describe rock art. Their model was able to find depictions in the art that were not previously noticed. This paves the idea that CNNs can be used to efficiently analyze archeological data and be used to find discoveries.

The classification systems used in the research mentioned above use predefined typologies identified by modern researchers. And while these systems may still be useful, other systems can be and should be developed that capture a more natural approach to categorizing ceramics such as one that focuses just on what it looks like. Where a vessel was made may matter less to someone than what kind of design is on it. The pottery can then be analyzed based on its appearance. Sometimes where a ceramic came from might be the most meaningful aspect of it to someone, sometimes the most meaningful aspect is its material, and sometimes the most meaningful aspect is simply what it looks like. The users of these ceramics may have cared about many other factors, but it is undeniable the importance of the appearance of pottery. Therefore a classification system that groups up artifacts based on their visual similarity may be of use. Furthermore, to allow for efficient processing of vast amounts of data, beyond what humans can do, leveraging the effectiveness of CNNs would be imperative. This would allow for a method to construct a computationally grounded typology system on the basis of ceramics' visual similarity. By creating a new data driven classification of one's data, they may also be able to find undiscovered or justify known correlations in their data. In this paper, both this method and an example of it in use will be presented.

Methodology

A workflow diagram for the proposed method to generate a new typology can be found in figure 1. This workflow describes how one can take a dataset of images and find a way to categorize these images efficiently and meaningfully. The overall procedure is to first process the images into high dimensional vectors, then find an appropriate way to cluster the data into pottery types of this new system. I used a dataset of Chinese Neolithic pottery put together by Zhao (2023).

VGG19

First, the images are given to a CNN model. Figure 1 and I both use a VGG19 architecture developed by the Visual Geometry Group (2015) but other architectures could still work. The model used should be pretrained on a large dataset so that it has a prior "intuition" of what things may look like. The reason this is helpful is so that visually similar images get transformed to data points that are close to each other and visually divergent images get transformed to points far from each other. This allows clustering algorithms to be used to categorize the data. I used a VGG19 model pretrained on ImageNet, a dataset of millions of labelled images.

PCA

After each image has been transformed to a high dimensional vector, one needs to reduce its dimensionality for efficiency and accuracy. Data with fewer dimensions is less computationally expensive to work with and equally if not more importantly, data with too many dimensions cannot be reliably clustered using algorithms like K-means. This is because when there are too many dimensions, all data points are pretty far from each other as each dimension contributes some amount to their euclidean distance. When everything is far from everything else, the distances then all become very similar and therefore it would be hard to identify meaningful clusters of data points. One can use principal component analysis (PCA), developed by Karl Pearson in 1901, to achieve this. With my dataset, I chose to reduce the dimensionality to 10, retaining 50% of the variance.

K-means

To cluster the data into categories of a new typology, this method suggests one use k-means clustering. K-means clustering is an algorithm developed by Lloyd (1982) that places one's data points into k clusters. This algorithm requires one to give it the number of clusters first before partitioning the data, therefore the method instructs one to test out this algorithm with a different number of clusters each time and then compare them all.

Silhouette and Davies Bouldin Index

The silhouette score (Rousseeuw 1987) and Davies Bouldin index (Davies and Bouldin 1979) are both metrics that evaluate clustering algorithms. The silhouette score gives values between -1 and +1 with higher values given to better clusterings and the Davies Bouldin index gives lower values to better clusterings with the lowest being 0. Using these two metrics together, one can find which of the tested clusterings should be their chosen model for their classification system. These two metrics together may suggest multiple ways to meaningfully partition one's data and therefore one has the option to pick which they would like to use. For example, this workflow may suggest one way to split one's data into 5 groups and another way to split it into 25 groups. Someone may reasonably think that 5 is too small or 25 is too big for their dataset and goals and could pick their preferred system.

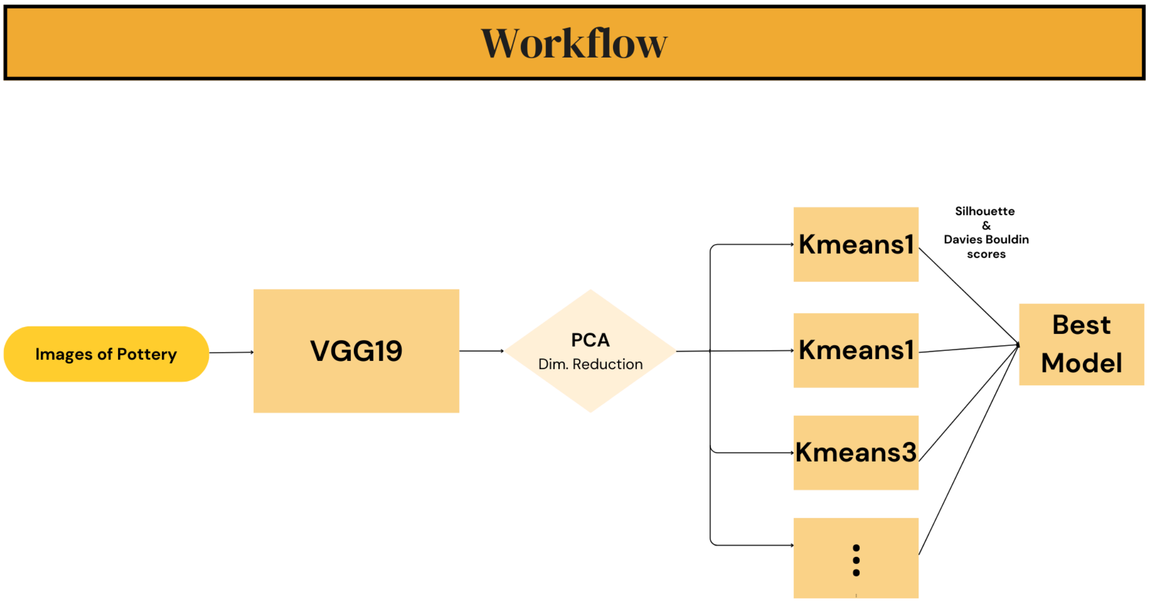

Figure 1. Workflow for the proposed method described in *Methodology*

Results

Developing the New Typology System

As mentioned above, I used this workflow to develop a system to classify Chinese Neolithic Pottery. The silhouette scores and Davies Bouldin indexes can be seen in figures 2 and 3 respectively.

The silhouette scores suggest that there can be two, three, or four clusters. The scores are relatively low, all being less than 0.29, however they are also all larger than 0.22 and would be expected to increase if the dimensionality were to be decreased. Higher dimensions lead to expectedly lower silhouette scores because the distances between two points in high-dimensional space tends to be pretty large. In addition to this, after four clusters, the silhouette score quickly drops. This suggests that four clusters would be a good number of clusters and would capture a bit more nuance in the data than two or three clusters because there are more categories to put a ceramic into (but not too many). The Davies Bouldin indexes suggest that there may be two or three clusters. These are two local minimums in the data with a few number of clusters.

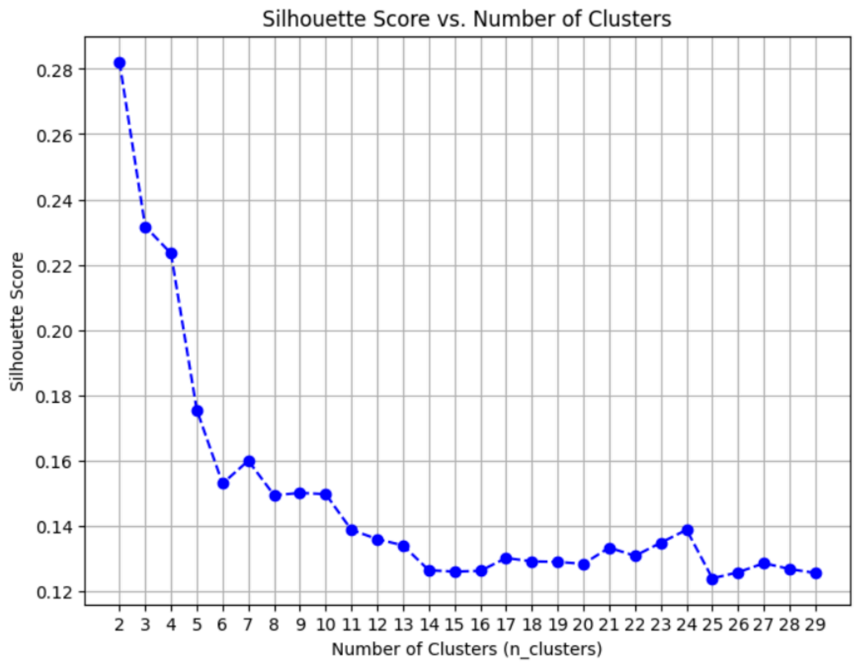

Figure 2. Silhouette scores for each of the k-means models initialized with a different number of clusters (n_clusters).

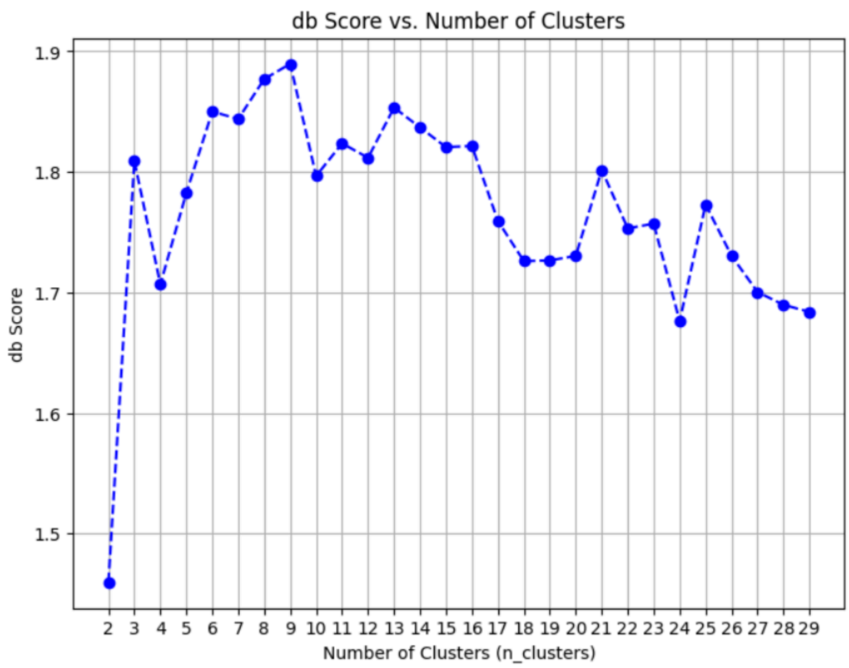

Figure 3. Davies Bouldin Indexes (or scores) for each of the k-means models initialized with a different number of clusters (n_clusters).

There are other local minimums on the graph, but they happen when the number of clusters is much higher. The other local minimums after n_clusters = 4 all are associated with low silhouette scores on the silhouette score graph so they are not suggested to be the clusters of good classification systems.

Interestingly there is a local minimum at 24 on the Davies Bouldin Index which matches with a low local maximum on the silhouette score graph, but it is still pretty low compared with the point at n_clusters = 2, 3, or 4.

From these two graphs together, one can argue that the k-means model initialized with 4 clusters would be the best for this new typology system. It has relatively good metrics and the number of types is large enough to discern differences between the ceramics.

Analyzing the New Typology System

From the results in the previous section, we can see that this new classification system has four types. In figure 4, I colored the data points of each cluster based on which cluster they're in and then projected the points onto 2D so that it can be visualized. There is some overlap between these clusters, but that is an artifact of the data being represented in 2D. In 10 dimensions, this overlap wouldn't exist.

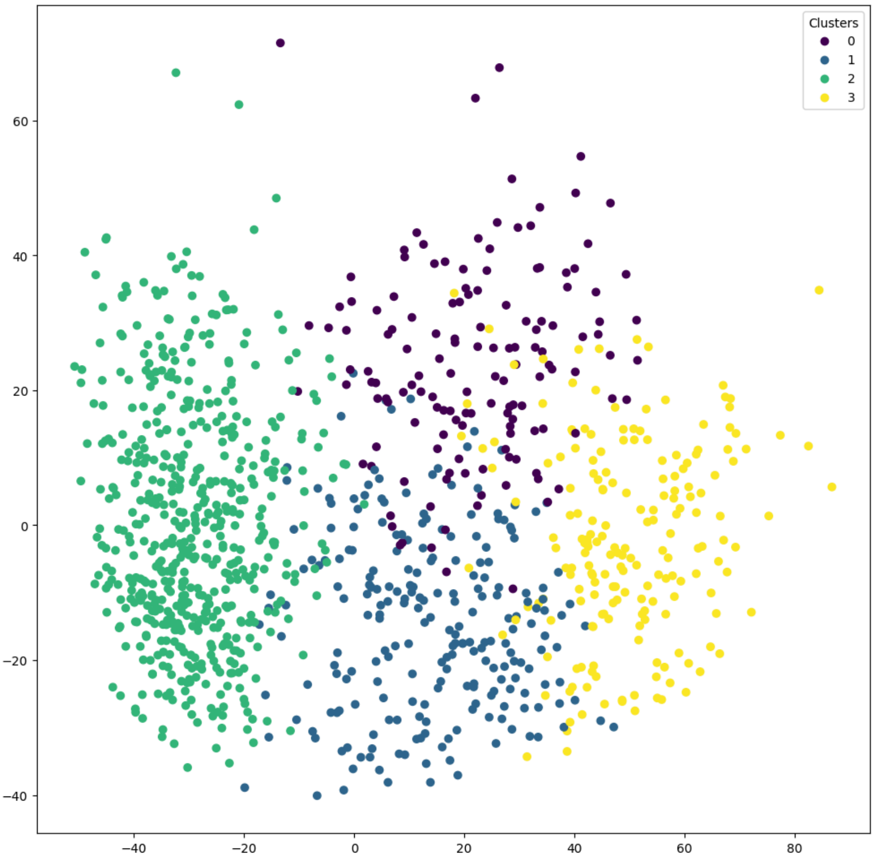

Figure 4. Data points from the Chinese Neolithic pottery dataset and colored based on what category/cluster they're in.

To show the use and effectiveness of this new typology system, I will analyze clusters 2 (left most) and 3 (right most). Further research can be done on this dataset, but the goal in this paper is to show this new method to generate a classification system and to prove its usefulness.

Figure 5 contains two images from cluster 2 and two images from cluster three. One can immediately see that cluster 2 likely contains many bowl-like pottery. It is also possible that the designs on these ceramics influenced what category they're in, but further research would need to be done to confirm that. Cluster 3, however, seems to have thin-necked pottery. While a display of two random images from these clusters doesn't tell us everything about what each cluster represents, it does show that it is likely this method generated a typology system that is meaningful to us.

Figure 5. Two randomly selected images of pottery from cluster 2 and from cluster 3. This shows that this classification system has likely identified categories that are meaningful to us.

The pottery comes from five archeological sites in China: Banpo, Banshan, Machang, Majiayao, and Miaodigou. Figure 6 shows what percent of the pottery in this dataset from each location are in clusters 2 and 3.

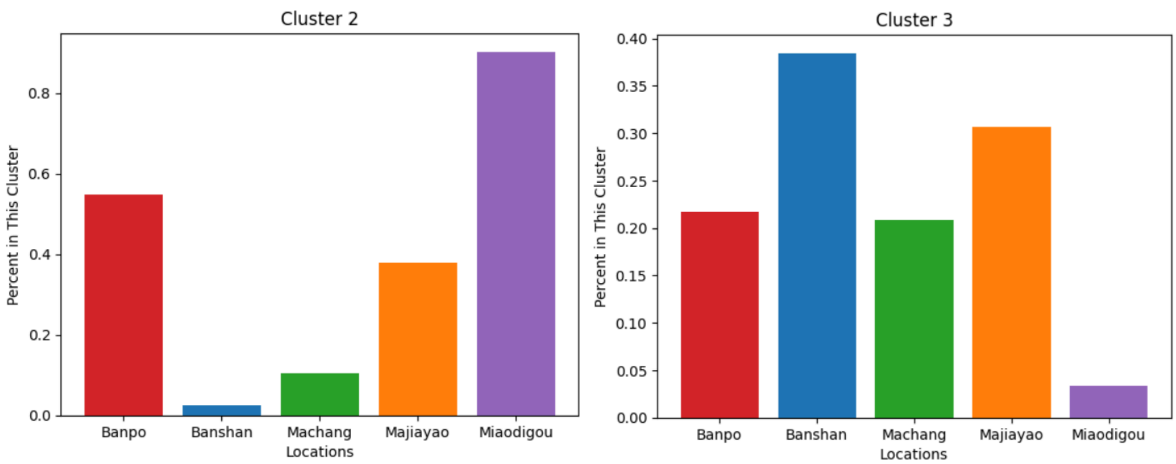

Figure 6. What percent of the pottery from each location is in cluster 2 or in cluster 3?

Discussion

Running the proposed methodology on a dataset of Chinese Neolithic pottery led to meaningful categorizations of the data and suggests that this new method can be of great use to many in this field.

Figure 6 shows that around 90% of Miaodigou pottery and almost 60% of Banpo pottery appeared in cluster 2 where fewer than 5% of Miaodigou pottery appeared in cluster 3 and a greater percentage of the pottery of other locations appeared in cluster 3. This suggests That there may be a connection between the Miaodigou culture and the Banpo culture and that they may differ from the Banshan, Machang, and Majiayao cultures. Liu and Chen (2012) write that indeed Banpo and Miaodigou cultures are two phases of the Yangshao culture and that the Banshan, Machang, and Majiayao cultures are three phases of a distinct culture that developed from the Yangshao culture through the Shilingxia culture that existed in the upper Yellow River region between 3980–3264 BCE. So this method of creating classification systems can also reveal information present in the data. Future exploration of this may reveal things yet to be present in the literature.

This method to generate new classification systems has two main flaws. One, it only takes into account appearances. While this can be very useful for many cases as I argued above, sometimes the provenance or the material, for example, may be very important to how something should be classified. Slight tweakings of this methodology would allow that, but further testing would need to be done to show how this can be done in an easy and effective way. Two, it requires a human to pick the best model. The metrics used in this methodology can give a researcher a very good idea of which model is the best, but it never says exactly which one to use. The benefit of this is that it allows the researcher to make sure the model is best fit for their goals. The problem may be that this isn't then 100% data driven and may suffer from human biases. Further research can be done to find a reliable way to test these clustering models that would determine which will be the most useful for researchers.

These results from testing this proposed methodology with Chinese Neolithic pottery have shown that this method works to find a meaningful way to classify archeological artifacts.

Code

Code for this paper can be found here:

https://github.com/BraydenKO/Artifact-Intelligence

Bibliography

Chetouani, A., et al. 2020. Classification of engraved pottery sherds mixing deep-learning features by compact bilinear pooling. *Pattern Recognition Letters, 131*: 1–7. DOI: https://doi.org/10.1016/j.patrec.2019.12.009

Davies, D. L., & Bouldin, D. W. 1979. A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-1*(2): 224–227. DOI: https://doi.org/10.1109/TPAMI.1979.4766909

Deng, J., et al. 2009. ImageNet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE: 248–255. DOI: https://doi.org/10.1109/CVPR.2009.5206848

Horn, C., et al. 2022. Artificial intelligence, 3D documentation, and rock art—Approaching and reflecting on the automation of identification and classification of rock art images. *Journal of Archaeological Method and Theory, 29*: 188–213. DOI: https://doi.org/10.1007/s10816-021-09518-6

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. 1998. Gradient-based learning applied to document recognition. *Proceedings of the IEEE, 86*(11): 2278–2324. DOI: https://doi.org/10.1109/5.726791

Liu, L., & Chen, X. 2012. *The archaeology of China: From the late Paleolithic to the early Bronze Age* (pp.). Cambridge University Press, pp. 190–232.

Lloyd, S. P. 1982. Least squares quantization in PCM. *IEEE Transactions on Information Theory, 28*(2): 129–137. DOI: https://doi.org/10.1109/TIT.1982.1056489

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. *Journal of Archaeological Science, 130*. DOI: https://doi.org/10.1016/j.jas.2021.105375

Pearson, K. 1901. LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 2*(11): 559–572. DOI: https://doi.org/10.1080/14786440109462720

Rousseeuw, P. J. 1987. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Computational and Applied Mathematics, 20*: 53–65. DOI: https://doi.org/10.1016/0377-0427(87)90125-7

Simonyan, K., & Zisserman, A. 2015. Very deep convolutional networks for large-scale image recognition. *arXiv*. https://arxiv.org/abs/1409.1556

Zhao, X. 2023. From classification to matching: A CNN-based approach for retrieving painted pottery images. *Mendeley Data, V1*. DOI: https://doi.org/10.1016/j.daach.2023.e00269