The Great Vase Race: Machine Learning Approaches for Image Recognition of Ancient Greek Vases

Data Science II: AC209B Final Project
Institute for Applied Computational Sciences (IACS)
Harvard University
Spring 2022

Austin Nguyen
austinnguyen@g.harvard.edu

Ivonne Martinez
ivonne_martinez@g.harvard.edu

David Harshbarger
davidharshbarger@g.harvard.edu

1 Introduction

Ancient Greek vases are arguably one of the greatest archaeological findings in history due to its ability to provide pictorial insight into one of the earliest human civilizations. Many of the Ancient Greek pottery is decorated with narrative paintings that visualized stories of popular myths and early Mediterranean life[1]. While historians have kept an impressive archive of the ancient Greek vases, it presents a unique challenge to categorize and catalog vases at scale. To better support historians and museum curators, our task is to develop a model that can extract unique characteristics of vases beyond hand-labeled meta-data provided by art historians.

Our goal is to develop a model for visual recognition of images to intelligently sort images into meaningful categories based on the objects and scenes depicted on the vases. Greek vases offer a uniform field in which representations are consistent and multiple visual expressions exist on individual objects and scene compositions. Images also contain captions or verbal tags that we can use to validate our approaches.

2 Data Preprocessing

For this project, we will use the Arms and Armor archive, a database of more than 100,000 images of ancient Greek vases indexed and organized. The Arms and Armor archive vases can be search by color, feature recognition, and verbal tags via a machine learning search algorithm. It is based on software developed by Jeff Steward at the Harvard Art Museums and deployed on an auto-scaling Postgres instance on Google Kubernetes Engine.

We extracted images using the Arms and Armor Archive API. Since the archive consist of 100,000 images, we extracted the first 50 pages for data exploration. The scraped data was then stored into a data frame for further formatting and cleaning. Figure 1 shows the data frame features, data types, and descriptions. We found that not all images were downloadable via the URL. We further processed to remove the images in which the API returned an error message.

Feature Name	Data Type	Description
id	int64	Vase ID (Multiple images per ID)
Source ID	object	Image Web Source ID
Vase Number	object	Vase Number within the Image
Fabric	object	Created by
Technique	object	Vase Art Technique
Shape Name	object	Vase Shape
Date	object	Data of Creating
Attributed To	object	Vase Attribution
Decoration	object	Imge Description on Vase
Publication Record	object	Data Corpus
Pleiades UR	object	Inflammation URL
Color	object	Vase Image Color
Provenance	object	Place of Origin
LIMC ID	object	Digital LIMC
LIMC Web	object	LIMC was Obtained
CAVI Collection	object	NaN
image	object	Image URL
Transcription	object	Image Transcription
Transcription Trimmed	object	Image Transcription Trimmed

Figure 1: Data Types Table

[1] https://mymodernmet.com/ancient-greek-pottery/

Images are inconsistent in its sizes. Image repository contain both close-ups of painted scenes and photographs of whole vases. This motivated the need to standardize dimensions. To do this, we fetched the dimensions of all images and found that the maximum dimension along the width or height is 720 pixels. We pad to ensure all images are 720 pixels by 720 pixels. Additionally, we collapsed dimensions to 256 pixels by 256 pixels to reduce their size to ease computational load. Beyond height and width, we also standardized all images to have 3 color channels (RGB).

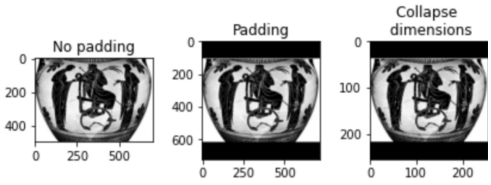

Figure 2: (a) No padding, (b) padding (720px), (c) reduced dimension (256)

Attached with each image is a useful set of metadata, containing information such as estimated date of origin (in windows of 50 or more years), place of origin (including modern and historical placenames), technique used (black-figure or red-figure), color (a separate metric from technique), shape, and – importantly – a short textual description of the scene displayed on the vase.

Some additional preprocessing is required to make full use of the metadata, typically by aggregating subcategories into larger groups. Places of origin smaller than countries were aggregated to the country level, and textual descriptions which included plural nouns were converted to their singular form so as to avoid splitting counts. Additionally, common words with little meaning, such as articles and prepositions, were removed from descriptions.

3 Exploratory Data Analysis

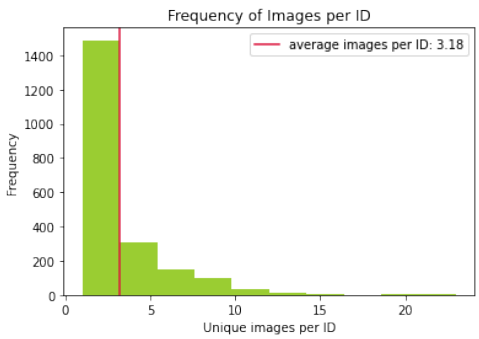

Figure 3: Images associated per unique ID

We fetched 6,696 unique images from the first 50 pages of the Arms and Armor Archives API. Due to storage and memory constraints, we worked with this initial seed list of images for the purposes of our EDA, but we plan to fetch a larger set of images downstream. Among the 6,696 images, we find that this maps to only 2,108 unique objects. This means that one museum object can have multiple images related to it. We find that the average images per ID is approximately 3.18. Thus, we can expect a single vase to be represented from different angles, crop sizes, and perspectives and we will be mindful of this later prevent leakage between our train-test splits.

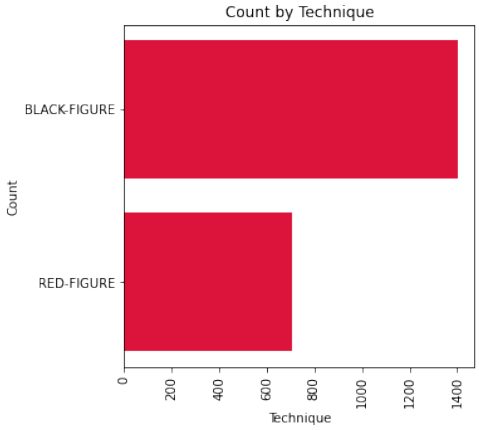

Figure 4: Frequency by technique

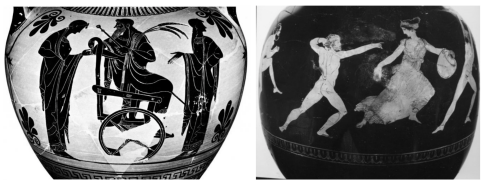

Figure 5: (a) Black-figure example (b) Red-figure example

We examine vases by technique. Black-figure vases outnumber red-figure vases, with roughly 66.6% vases are categorized as black-figure whereas 33.3% are categorized as red figures. Between the 6th and 4th century B.C.E, black-figure and red-figure techniques were used in Athens to decorate fine pottery. Invented in the city of Corinth around 700 B.C.E, black-figures is the older of the two styles and features black silhouettes set against colored clay [2]. In contrast, red-figure technique was invented around 530 B.C.E and feature red figure silhouettes on a black background. In black-figure vase painting, motifs were applied with a slip that

[2]Met Museum. "Athenian Vase Painting: Black- and Red-Figure Techniques." `https://www.metmuseum.org/toah/hd/vase/hd_vase.htm`

turned black during firing. In contrast, decorative motifs on red-figure vases remained the color of clay, inverting the background to be black. Based on this distribution of vases, the Arms and Armor collection appears to have many more black-figures.

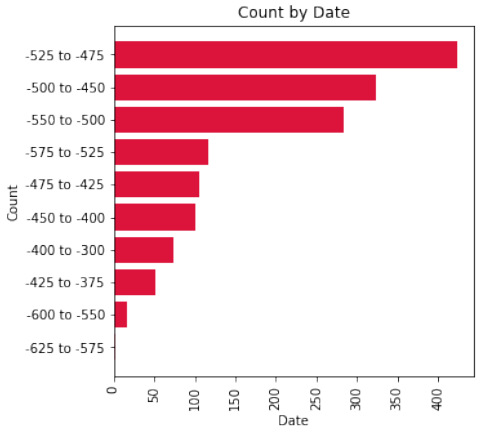

Figure 6: Frequency by date

We also examined the distribution of vases by date. 28% of vases were created between 525-475 B.C.E., 21.5% of vases were created 500-450 B.C.E, and 18.9% of vases were made 550-500 B.C.E. The time period aligns with our understanding of when black-figure paintings (around 700 B.C.E.) and red-figure paintings that soon followed.

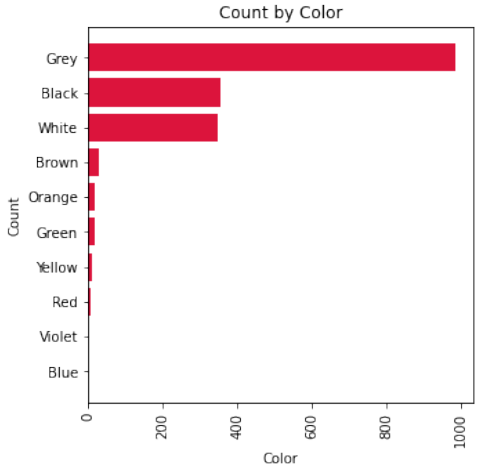

Figure 7: Frequency by color

We also examined vases by color. Gray comprised 55% of vases, black comprised 20% of vases, and white comprised 19.5% of vases. Color is different from painting style as it is possible to have a black-figure or red-figure on different colors of vases. We can think of color as a dimension that is orthogonal to painting technique. We find that the vast majority of vases were painted on gray vases, followed by white and then black. Moreover, the ratios of colors represented are relatively stable over time, unlike other features.

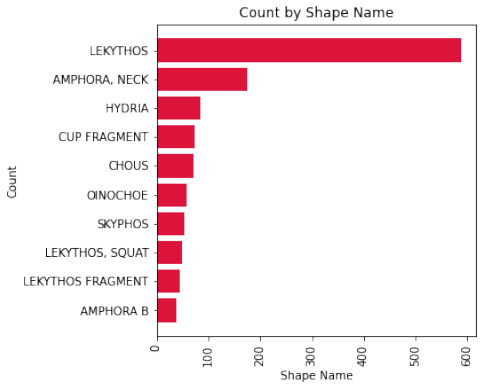

Figure 8: Frequency by shape name

We examined shape as a dimension. Lekythos comprised 27.8% of vases, Amphora, Neck comprised 8.3% of vases, and Hydria comprised 3.9% of vases. According to the Getty Museum[3], the Lekythos shape is characterized by a tall flask that is used to hold oil and in funeral rituals. The Amphora shape is characterized by two-handled storage jars that hold oil, wine, mill, or grain. The Hydria shape is used to carry water and typically made of bronze or pottery with three handles, two of which are used for carrying and one of which is used for pouring. This demonstrates that our vases are distinguished by function that dictate their shape which can may be useful features are model could learn (i.e. height of vase and/or number of handles).

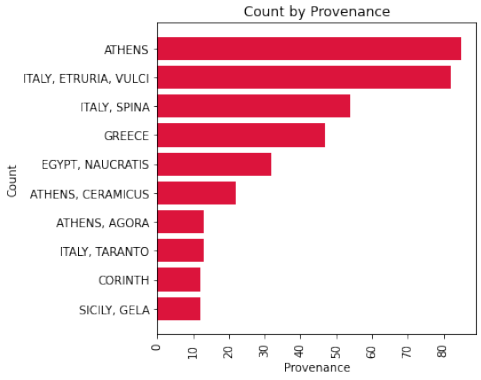

Figure 9: Frequency by provenance

Then, we examined the vases based on provenance which is the place of origin or earliest known history of a vase. 13.7% of vases are from Athens, 13.2% are from Italy, Etruria, and 8.7% are from Greece at-large. We find that most of the vases are from Athens followed by regions in Italy. This is consistent with our understanding of the dataset as

[3]Getty Museum. `https://www.getty.edu/education/teachers/classroom_resources/curricula/mythology/downloads/worksheet01_02.pdf`

most of the vases we would expect to come from Greece.

Next, we moved on to examining relationships between the features in the metadata, which will later be used to improve the interpretability of our models. First, we plot the use of black-figure technique and red-figure technique over time, and observe that neither technique spans the range of dates in the collection. Black-figure technique was enormously popular during the early years of this collection of vases, but very few such vases were collected from years after 450 BC, when red-figure techniques comprised nearly the entirety of the collection. The precipitous decline in Black-technique vases after this point is due either to a historical shift, a gap in the collection, or an unrepresentative sample of the collection due to our using only the first 50 pages of results for initial EDA.

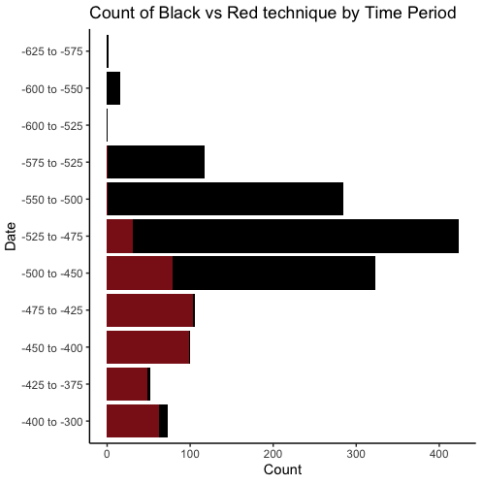

Figure 10: Technique used by era

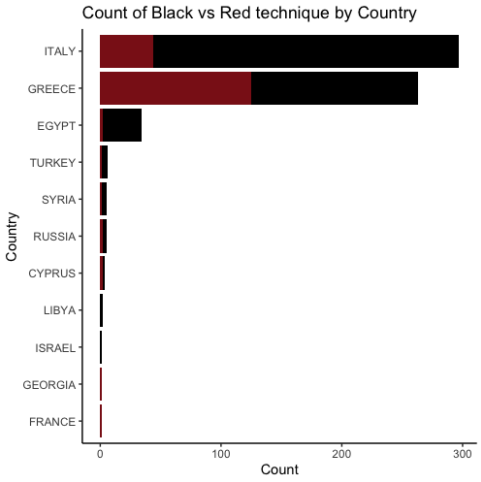

Figure 11: Technique used by era

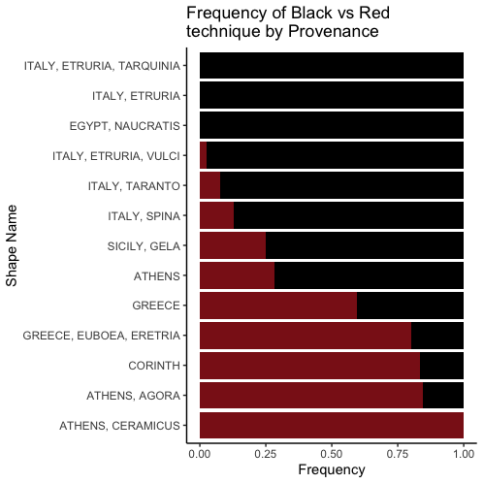

Figure 12: Technique used by era

Another notable shift in the dataset can be seen by considering the country of origin of older pieces in the collection as opposed to newer pieces. Italy and Greece comprise the vast majority of sources for the vases, but Italian vases were collected much earlier than Greek vases, overlapping mostly during only an estimated window of 75 years. Because this pattern is similar to that of the black-/red-technique divide, it is reasonable to question if technique is indicative of country of origin.

Fig 10 shows that the ratio of black to red techniques varies widely by country, and should not be assumed to be constant across space or time. Black-figure technique is much more common among Italian and Egyptian vases in the collection than Greek vases. Fig 11 shows this relationship in greater geographic detail.

One of the strongest indicators of technique is the specific shape of each vase. Among common shapes (20 or more vases in the sample), we observe a distribution of ratios, from the Olpe vase (exclusively painted with black-figure technique) to the Krater Bell vase (exclusively painted with red-figure technique) to many in-between. As both Olpe and Krater vases have a distinctive, classic shape, this relationship could inform the visual recognition of other correlated features. However, among most of the colors used on vases in the collection, the ratio of black- to red-figure technique roughly approximates the 2-to-1 share of the larger dataset, indicating that the color used is not likely to be one of the most important aspects of the visual appearance of vases for classification purposes. As a result, we may be able to perform classification reasonably well using a single-channel image, enhancing the efficiency of our model.

The final aspect of the metadata which proves useful for EDA is the textual description of each piece. Roughly 90% of images have some sort of text description, and after preprocessing we can see that the seven most commonly used (meaningful) words in the descriptions are WOMAN, YOUTH, DRAPED, WARRIOR, SEATED, CHARIOT, and MAN. However, these patterns of common words

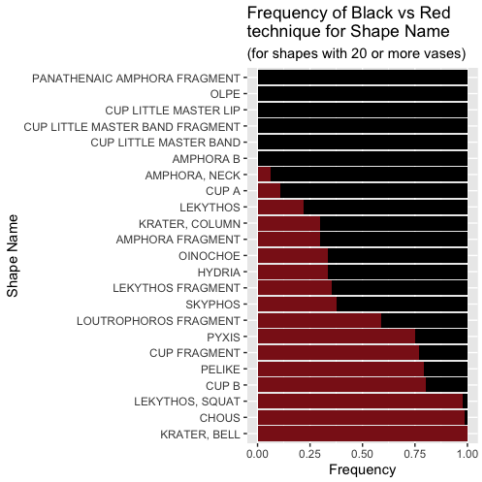

Figure 13: Technique used by era

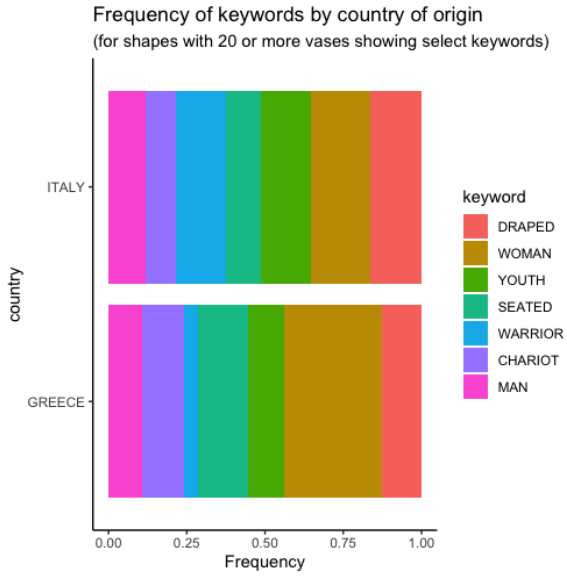

Figure 14: Technique used by era

do vary across other features like color and country of origin. While four of the seven most common words are shared among Greek and Italian scenes (WOMAN, DRAPED, YOUTH, and MAN), the words WARRIOR, DIONYSOS, and HERAKLES are much more common among Italian vases than Greek vases. Similarly, the words SEATED and CHARIOT are more common among Greek vases than Italian vases. The word FUNERARY appears on 21 different Greek vases, and is not used to describe any Italian vases.

Table 1: Seven most common keywords among Italian vases:

Keyword	Frequency
WOMAN	48
DRAPED	42
WARRIOR	41
YOUTH	41
DIONYSOS	31
MAN	30
HERAKLES	28

Table 2: Seven most common keywords among Greek vases:

Keyword	Frequency
WOMAN	87
SEATED	44
CHARIOT	37
DRAPED	35
YOUTH	33
MAN	30
FUNERARY	21

These cross-feature relationships will be used to guide our choice in network architecture when building a CNN for classification of new vases, as well as improve the interpretability of the model.

4 Methods

Based on our exploratory data analysis, we believe it would be most helpful to augment this textual information. Museum staff are currently able

While we are given metadata on the images via Arms and Armor API, our goal is to extract information from the images beyond the metadata. Museums currently have the power to sort images based on the text data and attributes hand-labeled by museum staff. To add value, we want to extract meaningful data from the images itself with which to sort in service of augmenting the categories already provided by the meta-data. For example, museum staff already have images labeled with "shield" and "spear". We hope to add to the list of categories. We also want a solution that can scale well with the numerous images in the catalog. We plan to do that by using a three-stage approach:

1. Convolutional Autoencoder: Feed input data into convolutional autoencoder to produce lower dimensional representation

2. Unsupervised Learning: Feed lower dimensional representation into unsupervised learning approach

3. Vanilla CNN: Train a follow-up convolutional neural network with the labels acquired from

unsupervised learning to implement saliency maps and understand the spatial relationship of pixels that result in that label

From this three-layer approach, we hope to recommend subset of labels that are meaningful to help museum staff better categorize Greek vases.

5 Implementation Plan

To add value to museums, our plan is to build a model that reduces the dimensionality of images and carrying out unsupervised learning approaches such as K-Means in order to cluster images based on pixels in the image. We will extract the clustering from unsupervised learning clustering to do implement saliency maps such as Grad-CAM. With Grad-CAM, we will exploit the spatial information that is preserved through the convolutional layers to understand which parts of the input image were critical to the classification decision.

Since our goal is to find vases that belong to the same objects and scenes we are considering to implementing crop augmentation or saliency maps in order to create labels for the elements drawn within our vases. One interesting approach that we are considering to develop this project is Self-Supervised Saliency Detection With No Labels. Being able to implement Self-Supervised Saliency Detection With No Labels will allow museums to faster process and analyze vases with images automatically.

Further approaches include unsupervised learning approaches, namely the clustering methods of hierarchical modeling and DBSCAN. We have learned from the metadata associated with each image that hierarchies do exist (for example, Provenance to Country and Shape to "more-general-Shape"), and that correlations persist across features. We therefore expect that the visual information will reflect those distinctions, and respond well to unsupervised learning models. We predict that these clustering methods will uncover existing groups identifiable from data.

Additionally, because the metadata are not a complete representation of the image, DBSCAN may prove useful in identifying outliers or groups of outliers, which could potentially constitute a real but unspecified grouping. Additionally, it could be useful for identifying the small number of images in the dataset which are not real images of vases, but sketches.

DBSCAN can sometimes fail to generate meaningful clusters if the dataset is too sparse, but with many thousands of similar vases (and being photographed in mostly standardized ways), we do not expect this to be a problem. Instead, the challenge of employing unsupervised learning approaches to this task will be to ensure that the methods by which we construct groups – either through hierarchical modeling or DBSCAN – actually reflect the desired information of the vases. Examples of desired information are details in the decoration or the shapes of the ceramic, rather than quirks of photography (such as dimensions, lighting, or angle), which might instead uncover a group of images taken by the same photographer, for instance.

5.1 Machine Learning Pipeline

1. Preprocess images: Standardize, pad, resize
2. Convert x into tensor objects: $x \in \mathcal{R}^{256x256x3}$
3. Data augmentation on data to produce multiple versions of image and improve model robustness downstream
4. Feed input images into a convolutional autoencoder
5. Feed output of previous step into unsupervised learning
6. Obtain labels from unsupervised learning and feed the input images and labels into a second CNN
7. Implement saliency maps based on labels
8. Examine which parts of images are highlighted in saliency maps to validate clusters obtained from unsupervised learning approach
9. Examine which parts of images are highlighted in saliency maps based on meta data as a sanity check

6 Early results

As part of the first stage of our approach, we implemented a convolutional autoencoder (CAE) to reduce the dimensionality of images. CAEs are a variant of convolutional neural networks (CNNs) that are used as the tools for unsupervised learning of convolution filters. We use CAEs to do image reconstruction to minimize reconstruction errors and produce a lower dimensional representation of images that can be used for downstream unsupervised approaches to cluster vases.

Early results from the first-stage of approach shows promise. Our convolutional autoencoder is able to extract important features in the early parts of the architecture, namely in the encoder portion. However, we do lose information in the decoder portion. This finding suggests that we may use the output of an the intermediate encoder layers to continue with unsupervised learning approaches such

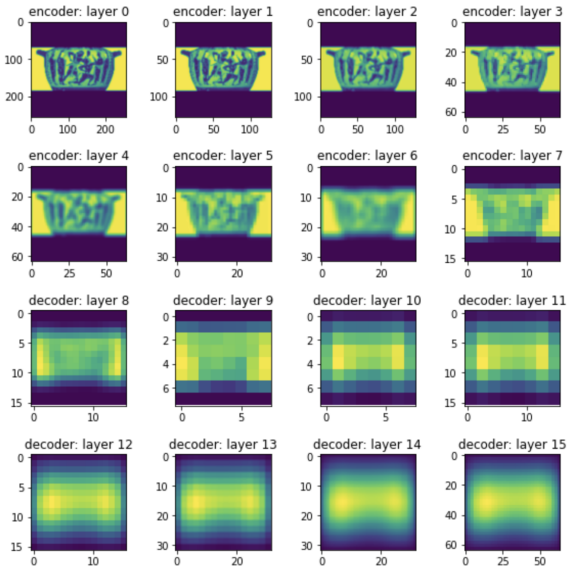

Figure 15: Visualizing features extracted from each layer of convolutional autoencoder

as K-means clustering and DBSCAN to produce meaningful clusters of similar vases.