Portfolio4 –Week 8-Image Clustering

Qiaoyu Wang

0 Introduction

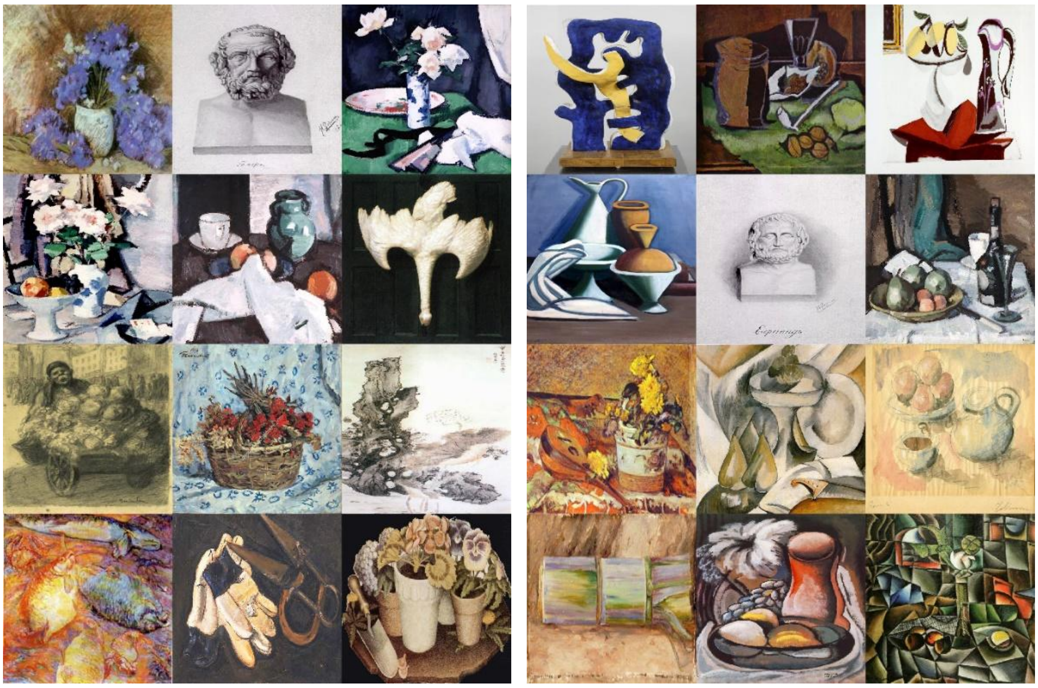
This week we recapped unsupervised learning and took a deeper look at image clustering. In class, we experimented with K-means image clustering, a simple yet powerful unsupervised machine learning algorithm. K-means clustering works by aggregating a set of data points together based on their similarities. We specify a target number k, which represents the number of centroids, and then allocate every data point to the nearest cluster, while minimizing the centroids' distance.

1 Image Clustering using lecture notebook

To begin, we downloaded nearly 800 images and used the ImageNet model to extract 1000 features from each image. Before clustering, we applied PCA to reduce the dimensions by extracting essential information and discarding less relevant features, thus facilitating a more straightforward analysis. And also we used the “elbow method” to identify the best k value and set it to 6. The resulting images depicted the clustering outcomes.

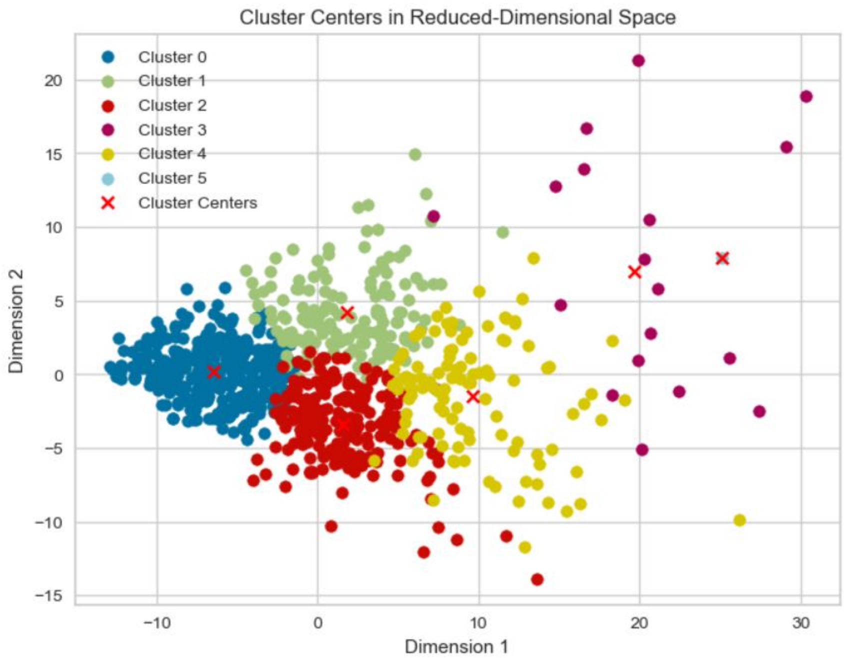
As a whole, the result of clustering was reasonable. It effectively grouped similar objects such as fruits, flowers, sculptures, and more. And also it successfully identified similarities in shape, color, posture, and composition, showcasing its efficacy. **However, when I got into closer observation, I found that two nearly identical ancient Greek images were placed in different categories within clusters 4 and 5.** To delve deeper into this discrepancy, I further investigated the images within these clusters to understand why these ancient Greek drawings were oddly divided.

The results were shown as following.



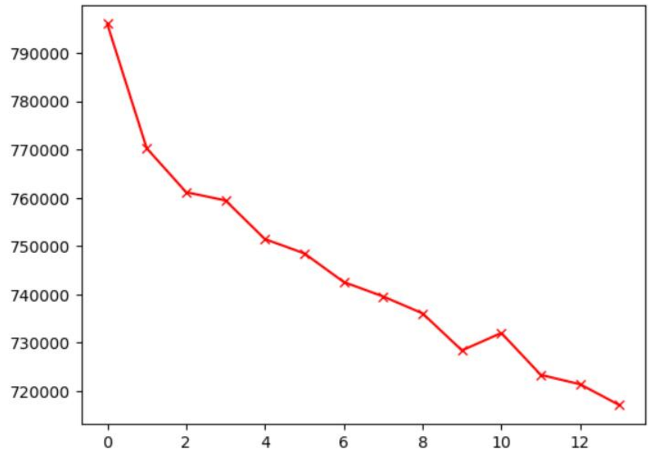
(Left: Images Nearest the Centre of Cluster 4; Right: Images Nearest the Centre of Cluster 5)

Upon closer examination, it showed that clusters 4 and 5 shared a similar overall appearance, which prompted an intriguing hypothesis: **Were the relative positions of these two clusters in the dimension plot figure closer than anticipated?** Let's delve into the dimension plot figure to explore this possibility further.

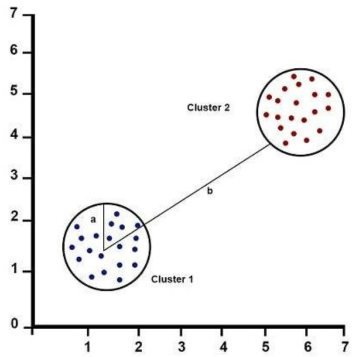


Now we encountered some intriguing observations. From the figure, **we hardly couldn't discover the position of Cluster 5, seemed like it just had one data point on the plot.** So why? **Was that because the value we put on k was not proper, which might not have effectively separated the data points into distinct clusters, leading to overlapping clusters or clusters that are difficult to visualize separately? Or the reduced-dimensional space created by PCA might not adequately capture the variation in the data, making it challenging to visualize the clusters effectively?**

Let's check the "elbow figure" which we used to find the best value of k. It seemed like point 6 was not a precise enough point to be considered as "elbow". **This beg another question: Is the elbow method the best way to find "k", especially there's not a clear elbow inflection point to identify the right "k"? Do we have other methods?**



I searched online and found another method called **Silhouette Score**, which could be used to study the separation distance between the resulting clusters. Let's try to validate the value of K using the Silhouette plot.



Silhouette scores for two clusters. | Image: Anmol Tomar

`Silhouette Score = (b-a)/max(a,b)`

Where:

- a = average intra-cluster distance, i.e the average distance between each point within a cluster.
- b = average inter-cluster distance i.e the average distance between all clusters.

- **1:** Points are perfectly assigned in a cluster and clusters are easily distinguishable.
- **0:** Clusters are overlapping.
- **-1:** Points are wrongly assigned in a cluster.

(Images from: Anmol Tomar, Aug. 02, 2023, Stop Using Elbow Method in K-Means Clustering: https://builtin.com/data-science/elbow-method)

Unfortunately, not the same as expected, I also got a poor Silhouette score. It showed that 5 out of 6 clusters exhibited below-average silhouette scores, and there were considerable fluctuations in the size of the silhouette plots. Besides, despite attempting various reduced dimensions, the results remained unsatisfactory.

Why didn't both the elbow figure and Silhouette score turn out as ideally expected?

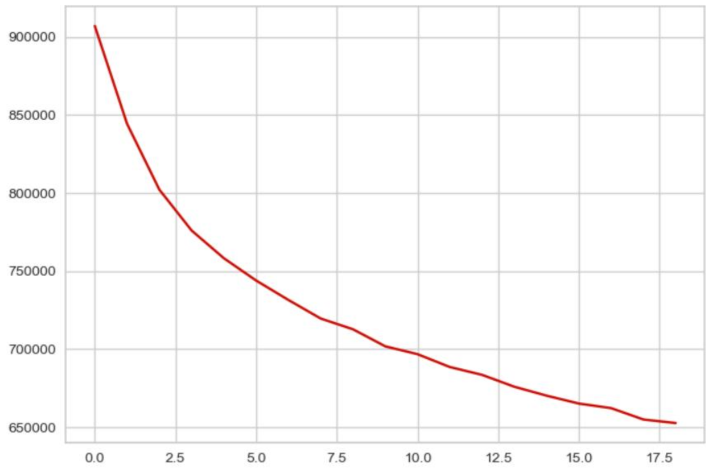
To understand this, we need to backtrack and examine how exactly the image clustering model operated. Clustering images is a multi-step process involving pre-processing the images, extracting features, clustering the images based on similarity, and evaluating the optimal number of clusters using a measure of goodness. This implies that for algorithms to cluster images effectively, they must first identify the features of each image, which we accomplished by utilizing the ImageNet model.

Let’s refer to the ImageNet class list (https://deeplearning.cms.waikato.ac.nz/user-guide/class-maps/IMAGENET/). If images are getting clustered together, it might be because they share some properties of these things. If not, perhaps our original dataset isn't sufficiently clear to be easily classified, even by a human. For example, distinguishing between pictures of cats and dogs is easy because they are concrete objects. However, when presented with numerous images of abstract things, like unidentifiable shapes requiring imagination to discern, clustering results will vary depending on individual perspectives.

Now it really makes sense to me. Therefore, I've decided to transition to a new dataset that fits the description I outlined above. I downloaded a new dataset from Pinterest, which comprises approximately 700 sculpture images, along with two datasets focused on ancient Greek images, totaling 300 images. All the images contain concrete and identifiable objects, which should ideally be classified clearly. Now, I'm trying to figure out how these pictures could be divided by K-means clustering and whether the ancient Greek images could be successfully distinguished among the sculpture pictures.

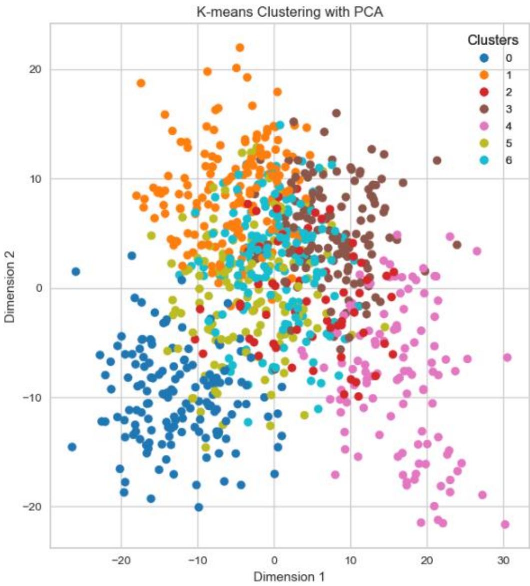
2 Image Clustering in new dataset

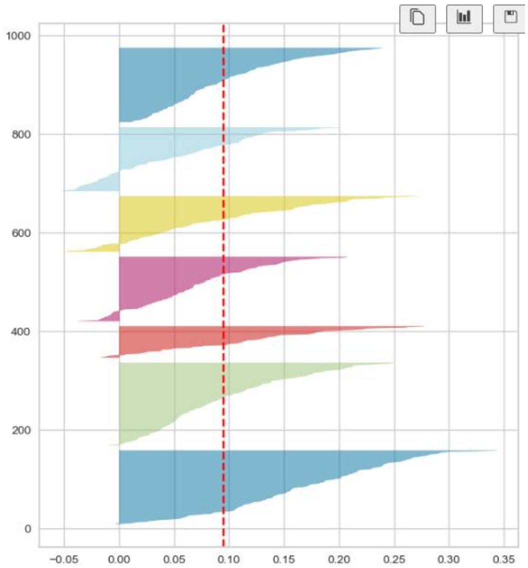
Firstly, I followed the same steps as before and generated an elbow plot of k. **It was encouraging to see that this time the figure showed a clear elbow, making it easy to identify.** Before I directly set k to its best value, which was 7 in this case, I also utilized the Silhouette method to assess the clustering quality.



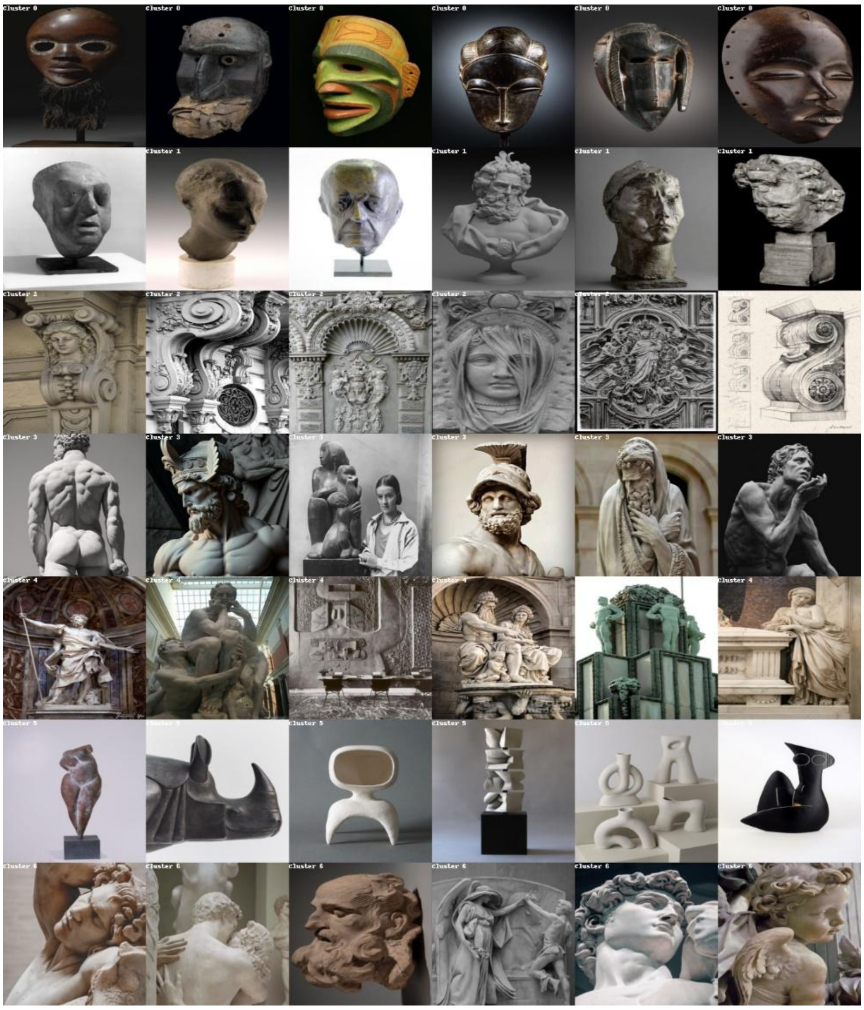
I was pleased to observe that all clusters exhibited above-average scores, and the fluctuations in each cluster's size were also acceptable. **However, despite these positive indicators, the overall Silhouette score remained relatively low, reaching only 0.09.**

Let's revisit the definition of the Silhouette method. It measures how similar an object is to its own cluster (cohesion) compared to other clusters (separation). **Therefore, if my original dataset exhibits considerable cohesion and similarity, it might not allow for sufficient separation between clusters, resulting in a low overall score.** This analogy resonates well with my dataset, which comprises numerous similar sculpture images, akin to dividing different sections within a single museum.





Let’s examine the representative pictures of each cluster. From my personal perspective, the images were clearly classified. We could see distinct clusters featuring specified head sculptures, column sculptures, body sculptures, and also simple shape sculptures. Looks like we can build a well-divided museum now!



3 Conclusion

Through this experiment, I gained a deeper understanding of each step involved in image clustering and the corresponding methods that can be employed. However, the most significant lesson I learned was the principle behind these methods: How exactly our computers identify new images, and how they categorize different images into meaningful clusters. It's fascinating to see how learning serves as a bridge, enabling us to form deeper connections with computers. Figuring out how computers interpret everything we humans have already been used to is an enlightening process. It allows me to challenge conventional thinking and reconstruct theories in our minds. It's truly enjoyable to unravel the mysteries of how computers perceive and categorize the world around us.