Image Classification of Ancient Egyptian Landmarks

Josh Sanchez

DIGS 30004 — Data Analysis II: Data Visualization and Machine Learning

6 March 2024

Introduction

The study of ancient cultures is often perceived by the public as a static and obsolete field with scholars pouring over texts or excavating ancient cities to uncover more about how the ancient world operated. However, the advent of machine learning, along with the rise of digital humanities, allows researchers of antiquity to uncover new information about a long gone era and how it could potentially relate to today. A common misconception is that the ancient world has no bearing on the modern; that its art and architecture, culture, politics, and lessons are completely separate from today's version. This could not be further from the truth. The numerous parallels between the United States and Ancient Greek politics, the Roman Empire's rise and fall, and Classical sculpture cannot be understated. The rise and collapse of the Roman Empire as well as the political system and turmoils of the ancient Greeks (specifically the ancient Athenians) compared to the United States can and does rely mostly on textual evidence and inferences. There is no quantifiable measure for which one can calculate the similarities between these systems and events. However, art and architecture, despite its intrinsic subjectivity, can be objectively compared between other ancient art and their modern counterparts. Typically, when attempting to discern information from unknown works of art from antiquity, other similar pieces are compared to it. When ascertaining if an ancient monument influenced the creation of modern works, both visual and textual evidence is used. The goal for this project is to build a machine learning model as a proof of concept that assesses the utility of automated visual recognition with regard to ancient monument comparison. The research question is being raised on how the model can be built to accurately and automatically identify if the monument is a structure or a sculpture.

Approach

The dataset in the final project was obtained from Kaggle, an online community for data scientists, and created by Kaggle user Marvy Ayman Halim. The dataset consists of twenty-two different ancient Egyptian monuments, totalling 3,576 images which are split up into folders that are labeled as either the monument's name (e.g. the bent pyramid for Snefru) or as the pharaoh that the sculpture depicts (e.g. Akhenaten). To avoid issues during the image preprocessing steps, one image that was stored as a .wav format was manually removed, decreasing the total number of images to 3,575. Additionally, a folder labeled "nodejs-crud" was manually removed as it did not pertain to the project.

The first step was to get the data into machine readable format. This was accomplished by preprocessing the folder and file names to make it lowercase and remove any spaces. These values are stored in a csv file and a Pandas dataframe. The dataframe has two columns, the first, labeled "image", stores the image filename, and the second, called "label", stores the folder name. Descriptions of the dataframe were produced to ensure the images were properly loaded (figure 1). Given the structure of the dataset, there was no need to remove missing information or drop unnecessary columns.

The next preprocessing step involved data transformations. Here, the data was resized and then grayscale. When working with any data type, it is important to consider balance when performing data cleaning and preprocessing. Specifically, clustering images without feature extraction heavily relies on balance in the images. Using images of varying sizes could cause the data to cluster around size rather than distribution of pixels. Another consideration when cleaning image data is preserving aspect ratio. While only having a small potential to change the clusters, aspect ratio could reveal a better model for image classification. Thus, in accordance with the research question, the resizing for this project was done by having two lists: one preserves aspect ratio while the second does not, in effort to determine what might provide better clusters, if any difference exists. To better visualize the difference that aspect ratio can make, contours (figures 2.1- 2.4) and histograms (figures 3.1-3.6) showing the distribution of pixel identities across the flattened image were produced. Three images from both the aspect preserved and non-aspect preserved lists were selected, the first image in the dataset (a statue of the pharaoh Akhenaten), a pyramid, and the last image in the dataset (a close up image of a face on the statue at the Great Temple of Ramses II). What those figures demonstrate is that the flattened images do have some variations in their distributions (seen in the first and last images), but no difference in the pyramid image. What this suggests, with regard to potential image clusters, is that certain images or types of images, such as specific angles of structures, may group identically across both aspect and non-aspect preserved modeling, but others may group differently in the two models. Now knowing that there is a non-zero probability of aspect ratio influencing the results of the clustering, the next step was to permanently flatten the images and store those values into their own columns in the dataframe. Figure 4 shows another description of the dataframe that was produced to ensure the values were properly stored.

Experiments and Results

Since the model is created using an unsupervised approach, there is no need to split the dataset into test and train sets, instead the whole set of flattened images can be used in the models. The first model used was *k*-means using two clusters (figure 5). This model showed some clustering of distinct groups, with one group closer to the bottom left, represented by the purple, and another group clustered from the middle to top left, shown in yellow. This visualization however, does not show clear clusters and is thus uninformative. Attempting to better visualize the clusters, Principal Component Analysis (PCA) was used to reduce the dimensionality of the images. 100 was the arbitrary number of components selected to reduce the dimensions. Using the result from the PCA reduction in the *k*-means clustering yielded a different view of the same data from figure 5. The distinction between the two groups was clearer, with a wider, larger yellow cluster on the left and a smaller, sparser purple cluster on the right (figure 6). Figure 7 demonstrates the attempt at using t-SNE to obtain clear and defined clusters from the images. There were some hints of clusters, shown by the darker blue spots scattered around the figure, but no clear clusters. In an endeavor to reveal the hints of clusters, PCA reduction was used in another t-SNE model, but there were no significant changes noted (figure 8). Unsurprisingly, considering the little variance between the histograms showing the flattened images preserving and not preserving aspect ratio, figures 9-12 illustrate how the results of the aspect preserved images modeling were almost identical to their non-aspect preserved counterparts.

While unable to be successfully reproduced, despite the underlying code remaining the same, it is important to note that one earlier model run of the t-SNE for non-aspect preserved images yielded a result much more akin to what was expected. Figure 13 shows the results of the model with what appears to be five, potentially six, clusters. Although incapable of providing information regarding what the model clustered on, the groupings appeared to be clustered based on similarity to being a sculpture or a structure. Structures total 1,320 images, or roughly 36 percent of the total dataset. Seeing that the left side is sparser and contains less data points than the right, one can reason that the left side represents structures. Sculptures total 2,256 images, or approximately 63 percent of the total dataset, and most probably make up the right half of the visualization. The top most cluster could be the Sphinx, as the sculpture has the third most images in the dataset with 345, or 9.6 percent of the entire dataset. Colossal in scale and defying most of the similarities found in the other sculptures as it is not a human, the model could have had a difficult time placing the Sphinx squarely in either category, thus placing it near the middle, but with a slight edge towards structure given is lack of sculpture similarities. The small bridge between the two halves of the figure could be landmarks such as temples or colossal statues attached to temples, like those found at the Great Temple of Ramses II. While these are statues with human traits, their placement directly in front of the temple could have potentially confused the model, thus leaving them somewhere in between sculpture and structure.

Criticism and Outlook

It is quite unfortunate that these attempts at an unsupervised approach to image classification to determine structure vs sculpture were unsuccessful. The core reasoning behind using an unsupervised approach for this particular project would be its adaptability. Envisioning this technology as a way to assist art historians and archaeologists, specifically those who work in the antique world, these researchers may not immediately know what comparanda could be used to verify an unknown monument.

This idea is perhaps best illustrated through an example. Imagine an archaeologist excavating a Roman site in Petra, Jordan and uncovers a fairly intact piece of art. Clearly, the archaeologist would know if it was a structure or sculpture, but what might not be known is what this monument is. Its provenance is certain and its date of creation might be contentious, but could be estimated. The monument's subject matter and style would be more difficult to determine and could always be debated, but these features could be interpreted with evidence from the rest of the excavation and what is known about the site at the time period that the monument was believed to be created. It is with these last two aspects, specifically with the monument's style that this project would help archaeologists. With a properly trained machine learning model, an image of the newly discovered monument could be added to the dataset to see where it clusters in accordance to other monuments from the culture.

Another example, this time within the world of art history, would be in ancient to modern monument comparison. Throughout the United States, there are hundreds, if not thousands, of monuments that echo that of the Classical Era. From the Lincoln Memorial in Washington, DC being modeled off of the Parthenon from Ancient Greece, to the Washington Square Park Arch in New York City being modeled off of the Arch of Titus from the Roman Empire, cities are filled with Neoclassical monuments. This project could aid art historians by more accurately and quickly identifying all possible matches. They would no longer need to spend large amounts of time sorting through the ancient monuments they believe a modern one echoes, instead utilizing a clustering model to visually see how closely related the Field Museum in Chicago is to the Erechtheion on the Acropolis in Athens.

As mentioned in the introduction, this project aimed to be a proof of concept looking that this idea could be accomplished. Given the time frame, what was believed to be a smaller, more achievable target was selected— cluster by sculpture or structure using an unsupervised approach, but that ultimately proved to be more complicated than originally anticipated. With no clearly defined groups, the model was unable to distinguish any meaningful characteristics from the flattened images. Currently, the model is looking at just the pixel distribution of the images, but this could be the source of the issue. If the images contain people within them, like a photo of tourists in the foreground with the Egyptian landmark in the background, the pixel distribution might be thrown off, making that image either biased to the sculpture cluster or as an outlier somewhere else on the plot. This behavior could have affected the *k*-means clustering, as *k*-means is sensitive to outliers. However, this sensitivity does not appear to have extremely impacted the model. The dataset itself could have been modified to better suit a *k*-means cluster using pixel distribution by only using images that contain just the landmark. This would eliminate photographs like those with tourists in them to more accurately align the image clusters. Furthermore, the issue of clustering based on pixel distribution could potentially be solved through feature extraction using a properly trained model like VGG16. Employing feature extraction, the model would assign more weight to the extracted features than the overall pixel distribution and would consequently cluster based on the similarity of the features in each image.

Another issue and critique with the project is the imperfect model fitting, and lack of variety in selecting models. While both *k*-means and PCA were used for linear models, and t-SNE, with and without PCA, was used for the non-linear attempt at clustering, other models may have worked, but were never implemented. Singular Value Decomposition (SVD) would decompose a matrix into a left, diagonal, and right singular matrix. Singular values would capture the variance of the data and singular vectors would represent the directions of greatest variance which could find the variance needed to classify into structure or sculpture. Non-negative Matrix Factorization (NMF) could have also been utilized. This technique decomposes a non-negative matrix into two non-negative matrices, which is useful when analyzing image data. Moreover, Linear Discriminant Analysis (LDA) could have been implemented either as its own technique or for dimensionality reduction in place of PCA. Projecting data onto a lower dimensional space where the separation between classes is best distinguished, LDA is useful for dimensionality reduction or feature extraction for classification, conceivably suggesting that feature extraction would have been a better approach.

Considering non-linear data, as evidenced by the clusters in figure 13, the project could have implemented techniques like autoencoders for feature extraction, dimensionality reduction, and anomaly detection. The encoders would have compressed the image data into lower-dimensional representations, capturing essential features, while the decoder attempts to reconstruct the original data. Deep Belief Networks (DBNs) could have also been utilized for either feature learning, dimensionality reduction, or image recognition. Kernel methods, such as Kernel PCA and Kernel *k*-means could have applied the techniques already implemented, but onto non-linear relationships. Lastly, a Convolutional Neural Network (CNN) could have been applied to learn complex relationships between input features, a widely used technique for image classification and object detection— core principles of this project.

Furthermore, attempts at classifying images based on labels, i.e. a supervised approach, could also be beneficial for the goals of this project. This would involve a change in the research question itself, which would now raise the question of how accurately the model could predict objects of its own type, i.e. given a training set of X images of Akhenaten, could it accurately predict a test set of Y images?

Moving slightly further back in the project timeline, the data cleaning is not perfect. When working with images, there is plenty of trial and error in the preprocessing stage. This then informs the model fitting, which can then lead to possible revisions to the data cleaning. For example, all of the images are grayscale in the model fittings, but what if color is essential to clustering these images? The potential bias at this point is assuming that because most landmarks are made of sandstone and limestone, thus making them beige, that grayscale would be inconsequential, and performed for efficiency purposes. Another issue could lie in the resizing. While the results of the models showed that aspect ratio was nearly irrelevant, the smaller image size could have resulted in a loss of quality necessary for image classification using the techniques from the project.

Conclusions

In conclusion, this project hoped to aid archaeologists and art historians, specifically those who study the ancient world, by modernizing and expediting the image classification process of unknown images. The specific research question looked to cluster images into either structures or sculptures in an attempt to provide a smaller scope for the project given the timeline. Following the research question, an ancient Egyptian image dataset was found and cleaned by making the text lowercase and replacing spaces with underscores. Preprocessing was then performed on the dataset by resizing (both to preserve and ignore aspect ratio) and grayscale to increase the efficiency of the model. With the cleaned data, an unsupervised approach using *k*-means clustering (with and without PCA), and t-SNE modeling (with and without PCA) were performed.

Through critiquing the data, the data preprocessing stage, and the modeling choices made, some bias in the processes was discovered and future improvements were realized. Considering the improvements that need to be made and potential biases that should be addressed, if this project were to continue in the future there is much that can be adjusted. To start, more modeling techniques should be used to see if there is a better fit for the data as represented after the current preprocessing steps. A mix of linear and non-linear would be used to arrive at the true best fit. Additionally, feature extraction should be attempted to determine if that is a more accurate or more efficient method of classifying images. This expanded project should then be replicated but with color preserved instead of grayscale images, and then again with a larger image resize. These new attempts will hopefully uncover the clusters searched for in the research question.

Figures

	image	label
0	15634726773_a8ac65d6ef_m_-_copy.jpg	akhenaten
1	19281291360_5a49331215_m.jpg	akhenaten
2	2906415757_50c2bc0414_m.jpg	akhenaten
3	41957529164_421e9f622f_m.jpg	akhenaten
4	4902788942_1c4ee56ede_m.jpg	akhenaten
5	7731634374_fe4e21a493_m.jpg	akhenaten
6	9711457465_051cf60521_n.jpg	akhenaten
7	a.1.jpg	akhenaten
8	a.10.jpg	akhenaten
9	a.11.jpg	akhenaten

Figure 1. Description of dataframe to ensure the images were properly loaded.

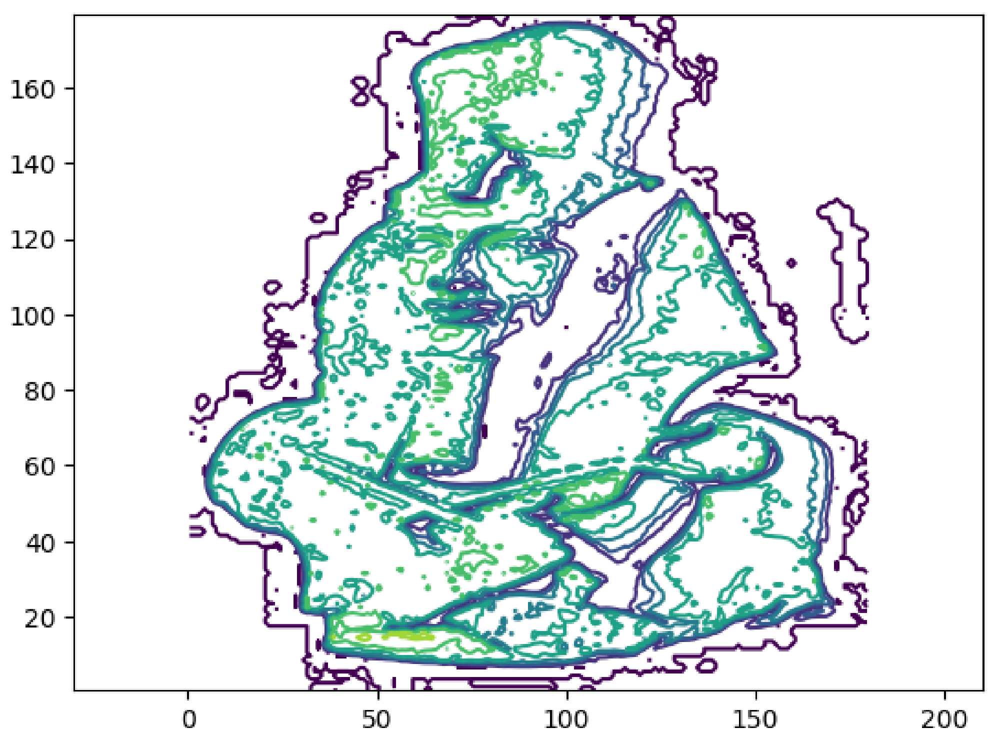

Figure 2.1. Contours of the statue of the pharaoh Akhenaten without preserving aspect ratio.

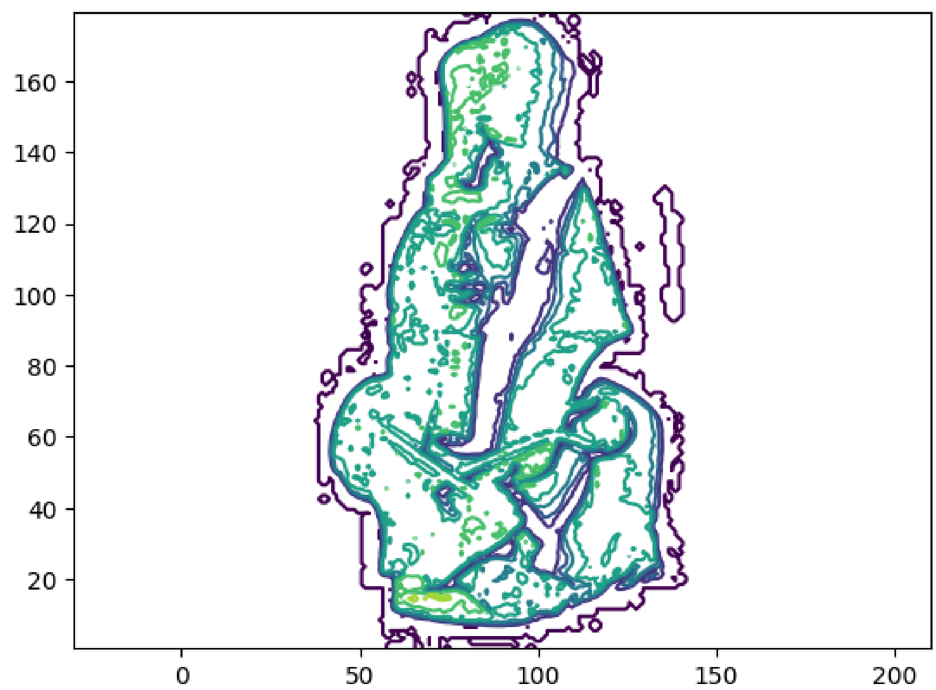

Figure 2.2. Contours of the statue of the pharaoh Akhenaten preserving aspect ratio.

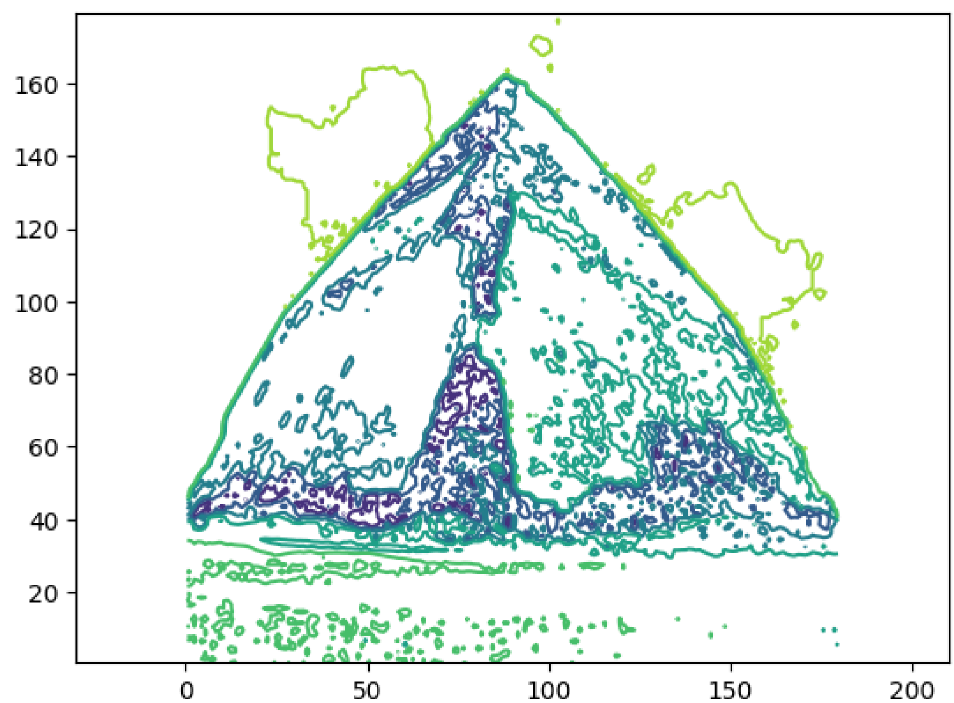

Figure 2.3. Contours of the Bent Pyramid for Snefru without preserving aspect ratio.

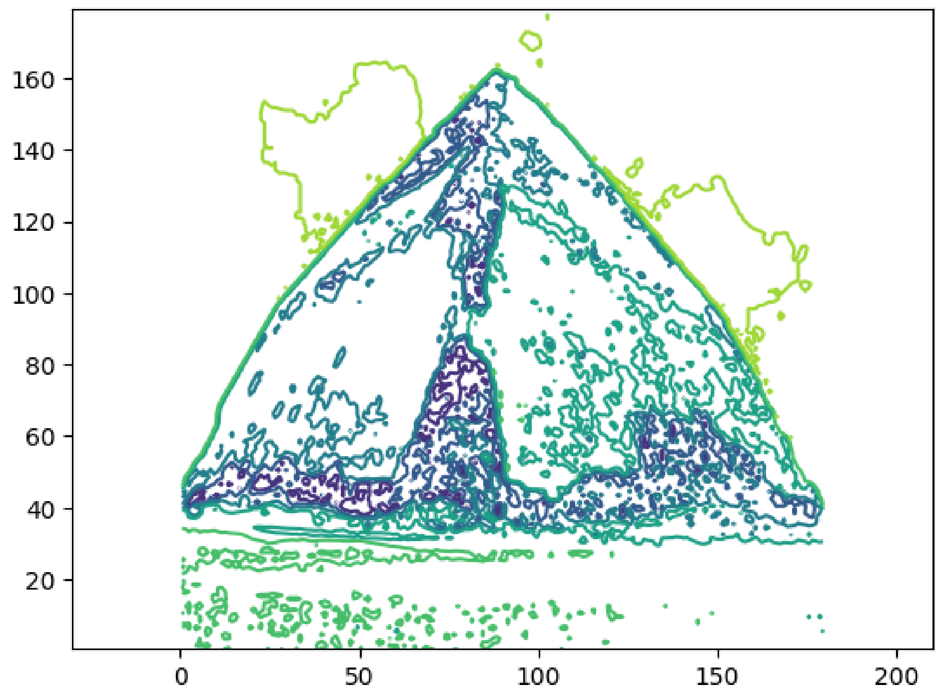

Figure 2.4. Contours of the Bent Pyramid for Snefru preserving aspect ratio.

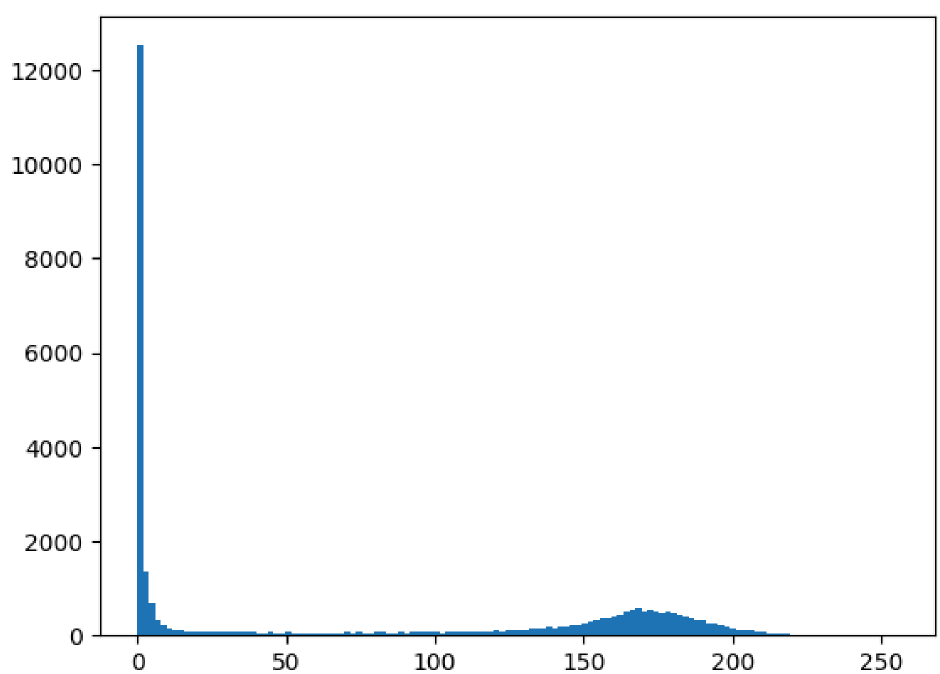

Figure 3.1. Histogram of the flattened image of the statue of the pharaoh Akhenaten without preserving aspect ratio.

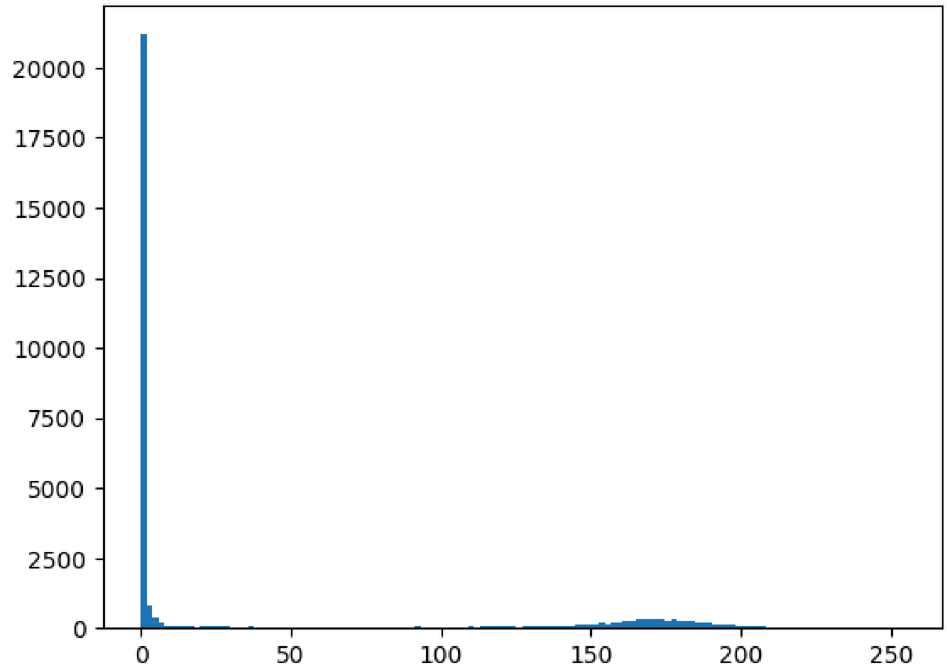

Figure 3.2. Histogram of the flattened image of statue of the pharaoh Akhenaten preserving aspect ratio.

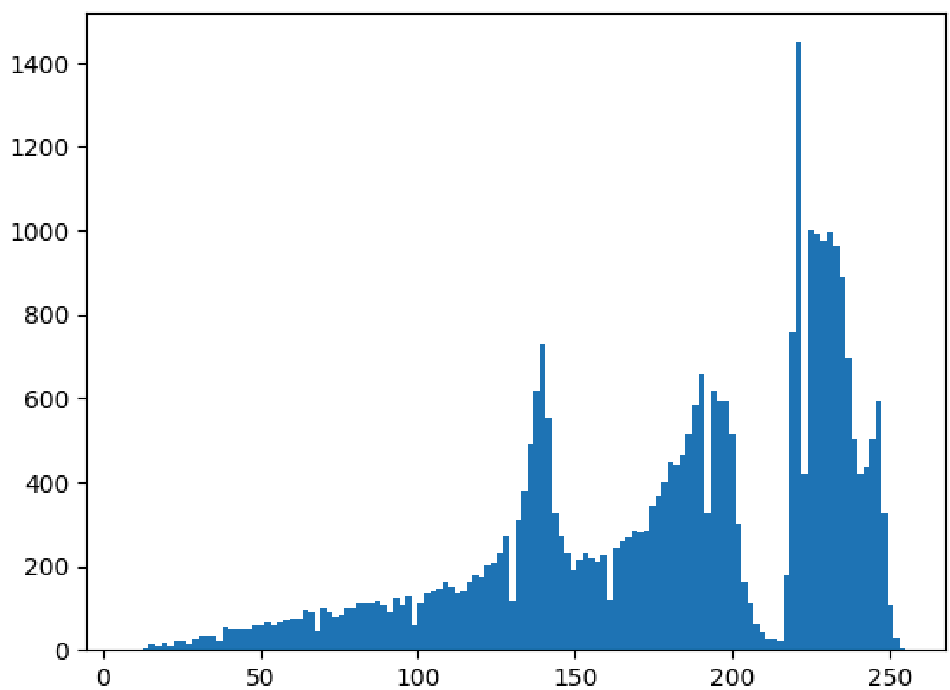

Figure 3.3. Histogram of the flattened image for the Bent Pyramid of Snefru without preserving aspect ratio.

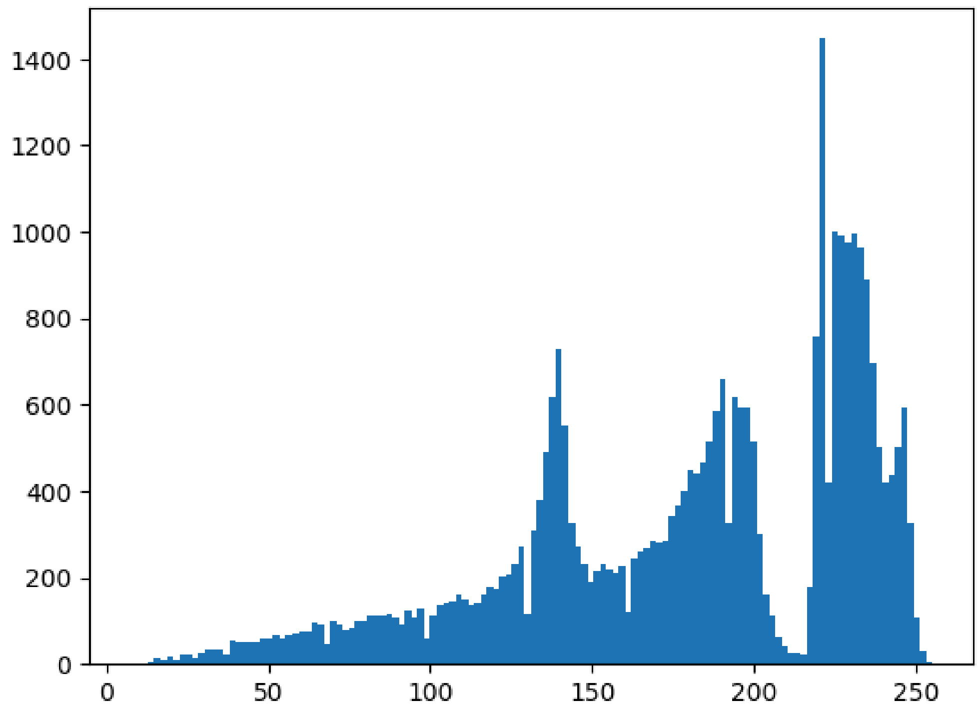

Figure 3.4. Histogram of the flattened image for the Bent Pyramid of Snefru preserving aspect ratio.

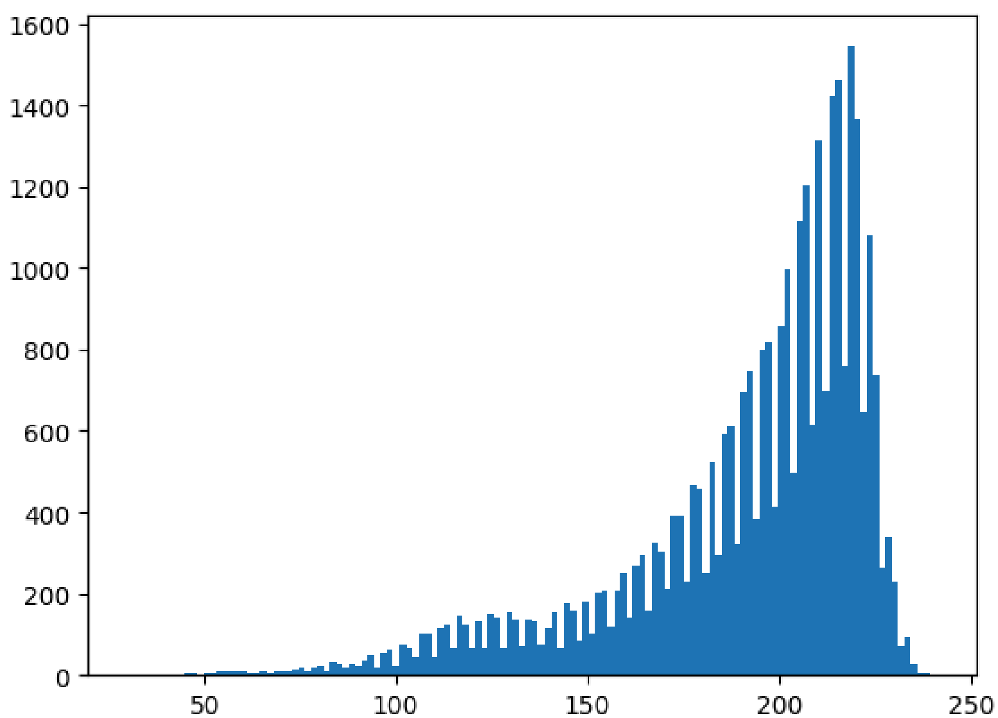

Figure 3.5. Histogram of the flattened image of the close up image of a face on the statue at the Great Temple of Ramses II in the dataset without preserving aspect ratio.

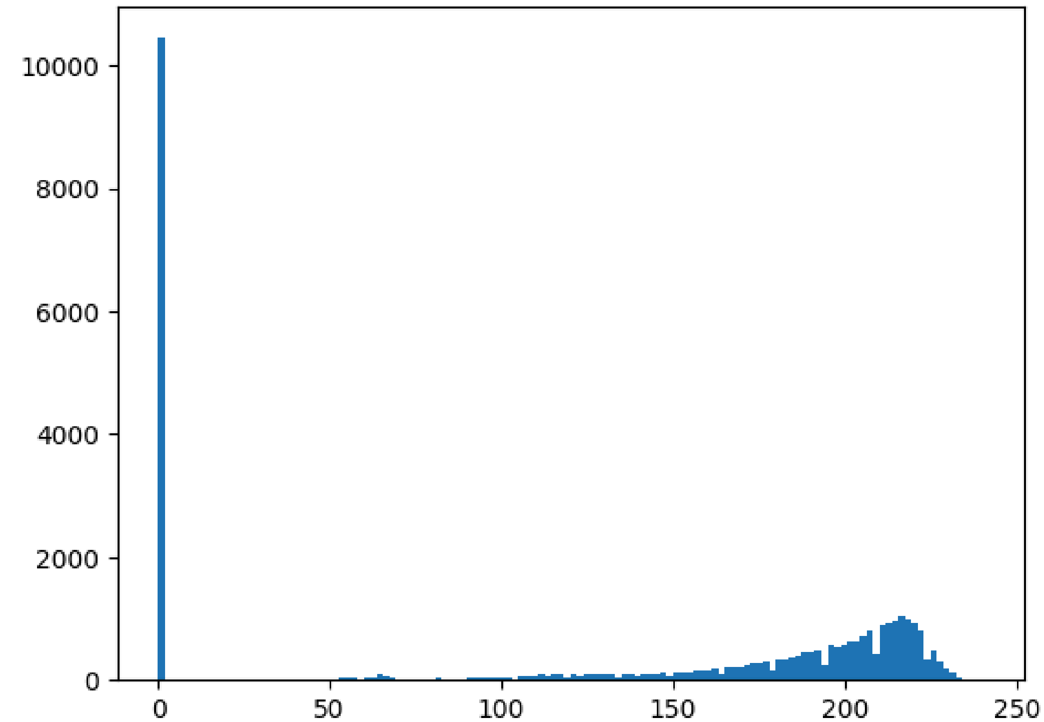

Figure 3.6. Histogram of the flattened image of the close up image of a face on the statue at the Great Temple of Ramses II in the dataset preserving aspect ratio.

	image	label	flattened_image	flattened_image_aspect_ratio
0	15634726773_a8ac65d6ef_m_-_copy.jpg	akhenaten	[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ...	[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ...
1	19281291360_5a49331215_m.jpg	akhenaten	[21, 21, 21, 21, 21, 22, 22, 22, 22, 22, 22, 2...	[21, 21, 21, 21, 21, 22, 22, 22, 22, 22, 22, 2...
2	2906415757_50c2bc0414_m.jpg	akhenaten	[63, 52, 45, 46, 46, 46, 46, 42, 42, 43, 43, 4...	[63, 52, 45, 46, 46, 46, 46, 42, 42, 43, 43, 4...
3	41957529164_421e9f622f_m.jpg	akhenaten	[255, 255, 255, 254, 255, 255, 255, 255, 255, ...	[255, 255, 255, 254, 255, 255, 255, 255, 255, ...
4	4902788942_1c4ee56ede_m.jpg	akhenaten	[244, 246, 247, 248, 248, 241, 246, 247, 247, ...	[244, 246, 247, 248, 248, 241, 246, 247, 247, ...

Figure 4. Description of the dataframe with the new flattened images.

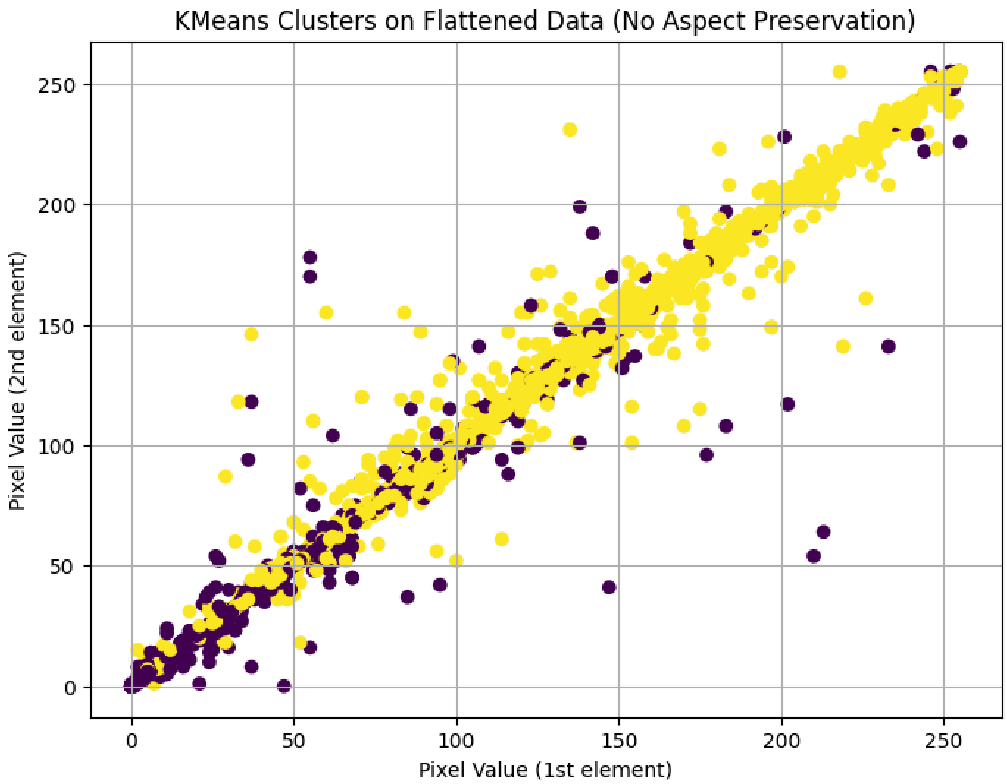

Figure 5. KMeans cluster on non-aspect preserved flattened images.

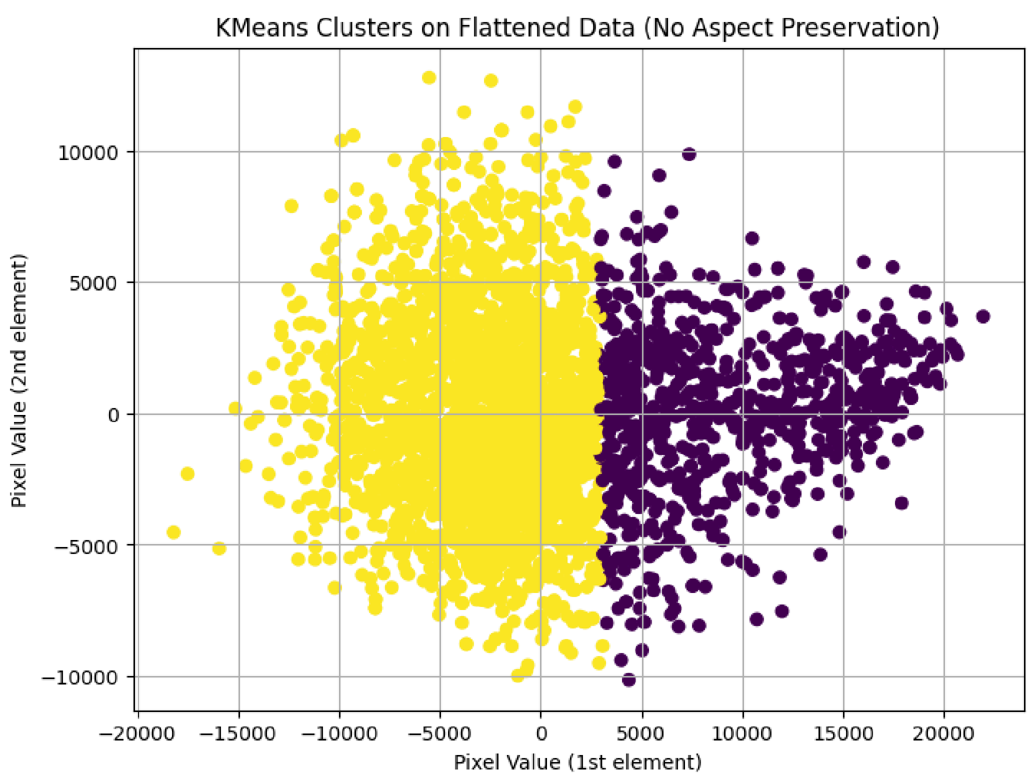

Figure 6. KMeans cluster on non-aspect preserved flattened images using PCA reduction.

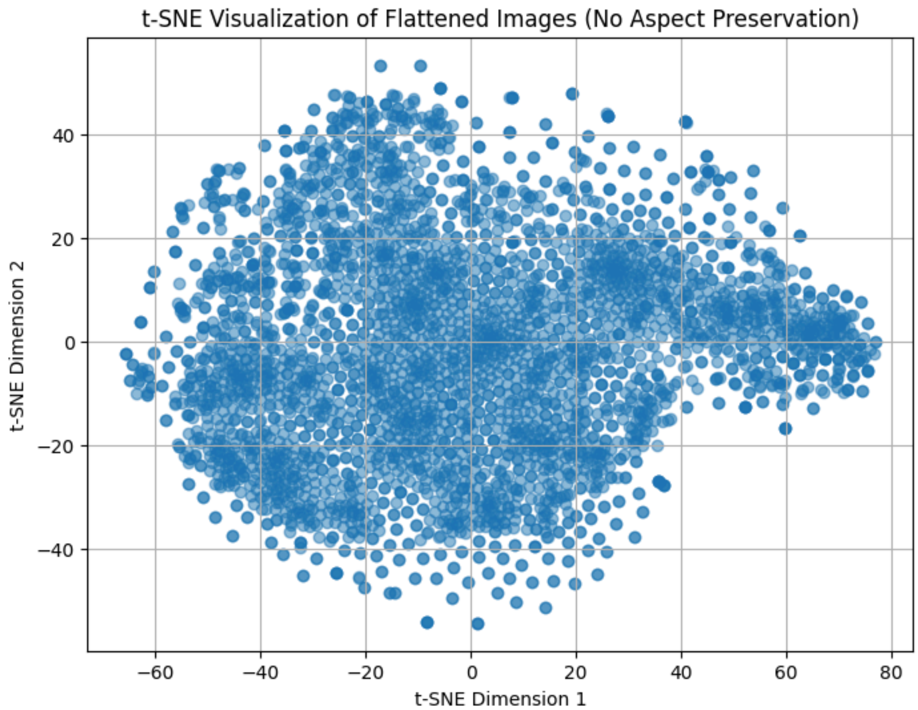

Figure 7. t-SNE on non-aspect preserved flattened images.

Figure 8. t-SNE on non-aspect preserved flattened images using PCA reduction.

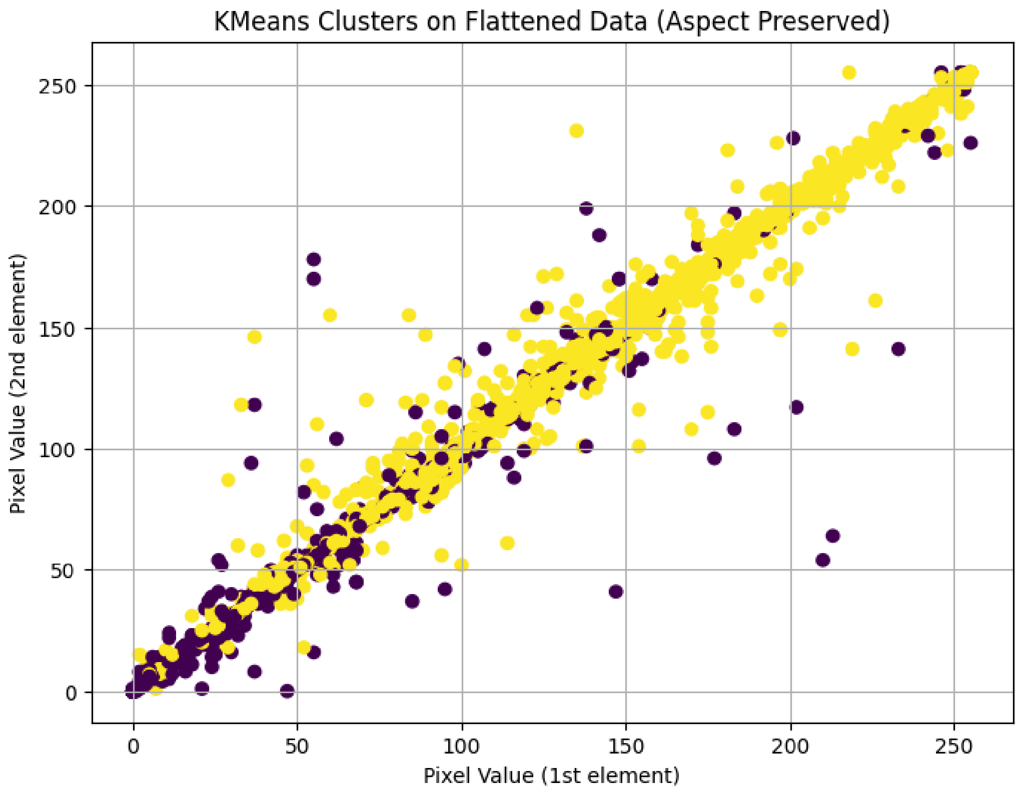

Figure 9. *K*-means clustering on aspect preserved flattened images.

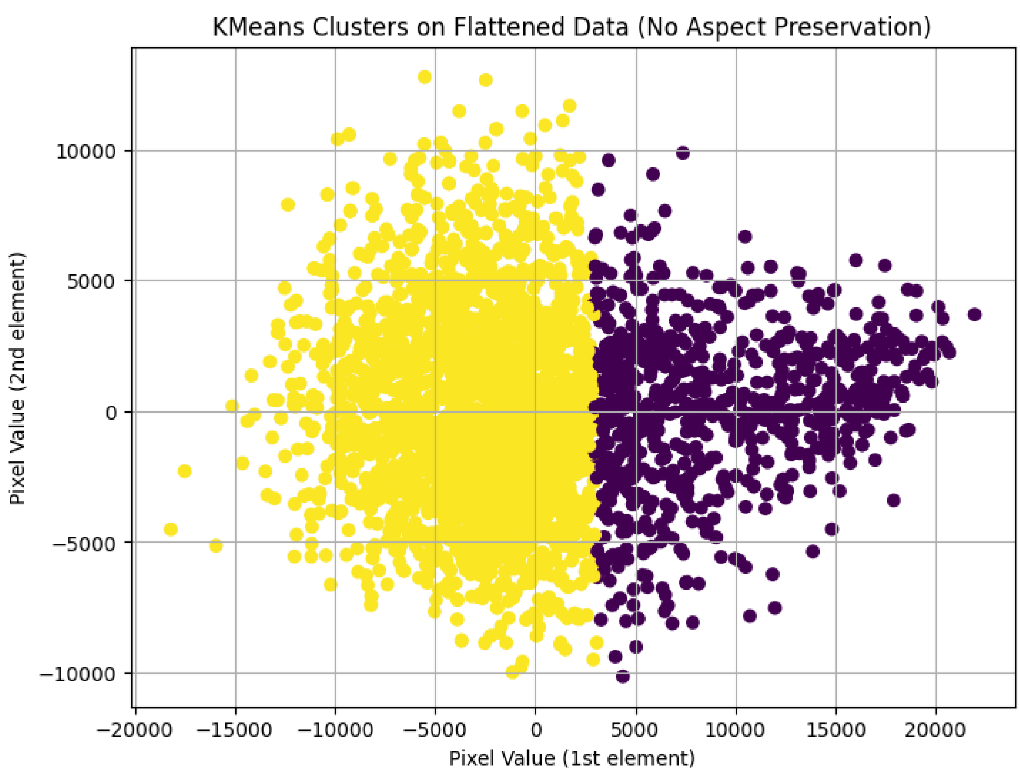

Figure 10. *K*-means clustering on aspect preserved flattened images using PCA reduction.

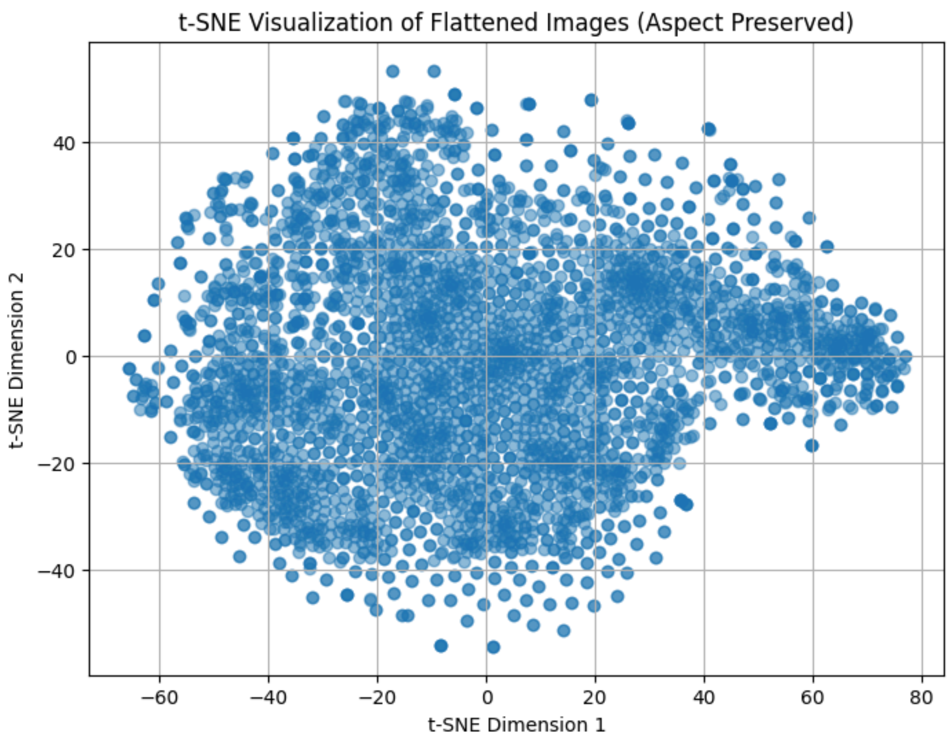

Figure 11. t-SNE on aspect preserved flattened images.

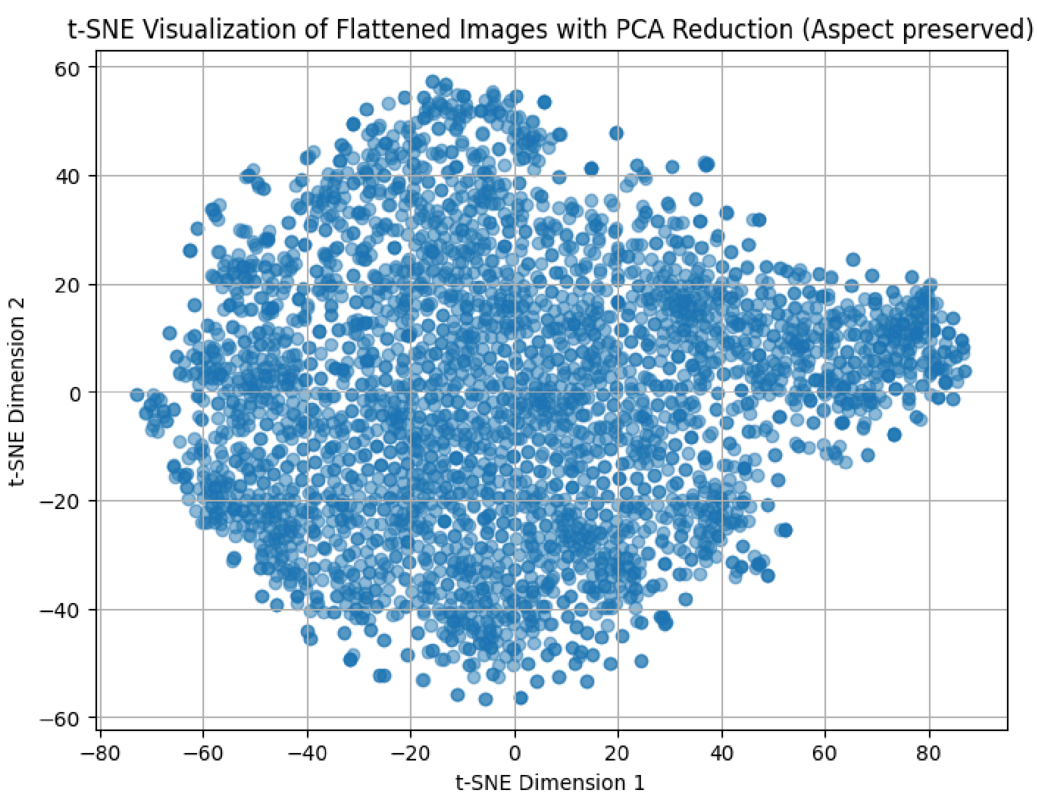

Figure 12. t-SNE on aspect preserved flattened images using PCA reduction.

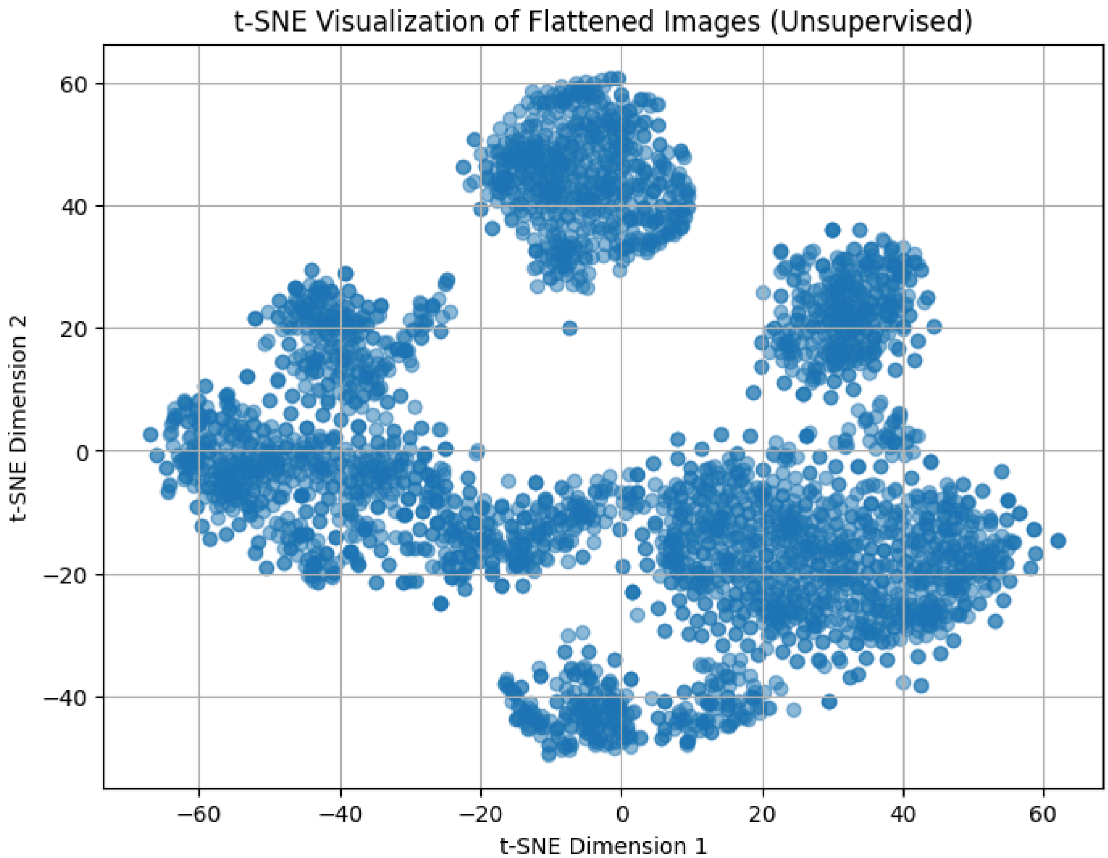

Figure 13. Early and unreproducible t-SNE visualization of flattened images without preserving aspect ratio.