# Towards the use of Adversarial Auto Encoders (AAE) to determine the cultural influence of ancient civilizations in Hellenistic Babylonian figurines.

Ishna Satyarth
*Lyle School of Engineering*
*Southern Methodist University*
Dallas, USA
isatyarth@smu.edu

Jesús Alejandro de Leon Moreno
*Lyle School of Engineering*
*Southern Methodist University*
Dallas, USA
aldelemo96@gmail.com

Yasmín Alejandra Femerling García
*Lyle School of Engineering*
*Southern Methodist University*
Dallas, USA
yfemerling@mail.smu.edu

***Abstract*—Images of sculptures from Achaemenid, Classical Greek and Hellenistic, Hellenistic Babylonyan, and Neo-Babylonian cultures from the Department of Art History of SMU, Dallas were encoded using an Adversarial Auto-Encoder into a latent 2D space to determine if Hellenistic Babylonyan art has cultural influence from Classical Greek, Neo-Babylonian, or Achaemenid art. Although the results are difficult to interpret, they can be used to help with the subjective resolution of art and historic research in the matter and to set ground to future work.**

***Index Terms*—AAE, art, style, figurines**

## I. Introduction

Historic art researchers have studied artifacts from different time periods and cultures, which can lead to a better understanding of said cultures. There are traits and characteristics that are identifiable and unique to certain civilizations, so influences from these can be found in art. By studying these influences, the interaction between different cultures and the way they merged can be learned from art. That is why it is important for historians and art researchers to determine the influence that art from a civilization had on another. Sometimes the task is complicated, because these characterization of traits and influences can be purely subjective. To research this problem, an Adversarial AutoEncoder (AAE) will be used to characterize art from different civilizations into a latent space. This could help researchers with the characterization of certain traits that could answer how influenced were newer civilizations by the art and culture from older civilizations.

To narrow the problem, a set of figurines from the Hellenistic Babylonian, Classical Greek and Hellenistic, Neo-Babylonian, and Achaemenid civilizations provided by SMU's Department of Art History will be studied. The research conducted by the department tries to identify the role of anthropomorphic figurines as agents of cross–cultural identity production and social negotiation in Hellenistic Babylonia. [1] To aid in this, an algorithm using an AAE will be used to represent the figurines in a 2D latent space. The Hellenistic Babylonian figurines can then be identified as being encoded in a space separate or close to the Neo-Babylonian, Classical Greek or Hellenistic, or Achaemenid figurines.

### *A. Historical Context*

It is also important to understand the interactions of the cultures to be analyzed, especially which civilizations invaded others, to help with the interpretation of the final results.

The Neo-Babylonian Empire, also called Second Babylonian Empire, existed from 626 BC to 539 BC. It was characterized by the rule of Nebuchadnezzar, bringing the golden age of the Empire, where it conquered Uruk and Nippur among other cities. In 539 BC, Babylonia was invaded and taken down by the Achaemenid king Cyprus the Great. The empire survived for centuries under the rule of the Achaemenid culture and later by the Hellenic Macedonian culture. [2]

The Greek culture had an enormous influence on the Mediterranean thanks to the military campaigns and invasions of Alexander the Great. This period was marked from the 5th to the 4th Century BC as the Classical Greek period. During 330 BC, Alexander the Great conquered Babylonia, starting the Hellenistic period. This started a widespread migration of Macedonian and Greek people. This Hellenistic period ended until the rise of the Roman Empire in 31 BC. The widespread migration, and cross-cultural interaction led to complex societies that are still being studied.

In both empires, the use of figurines was used for religious and artistic motives. During the Hellenistic period, figurines were used as charms to deter misfortune or to cast spells. During the Neo-Babylonian empire, terracota figurines were common. They were used as sacred objects and were also made as charms for magical protection or for offerings to deities in the temples. [3]

## II. Proposed Model

### *A. Adversarial AutoEncoder*

The problem at hand will involve not only objective computational processing, but also the subjective interpretation of said results. That is why the proposed model is an adversarial autoencoder, where a GAN performs variational inference by

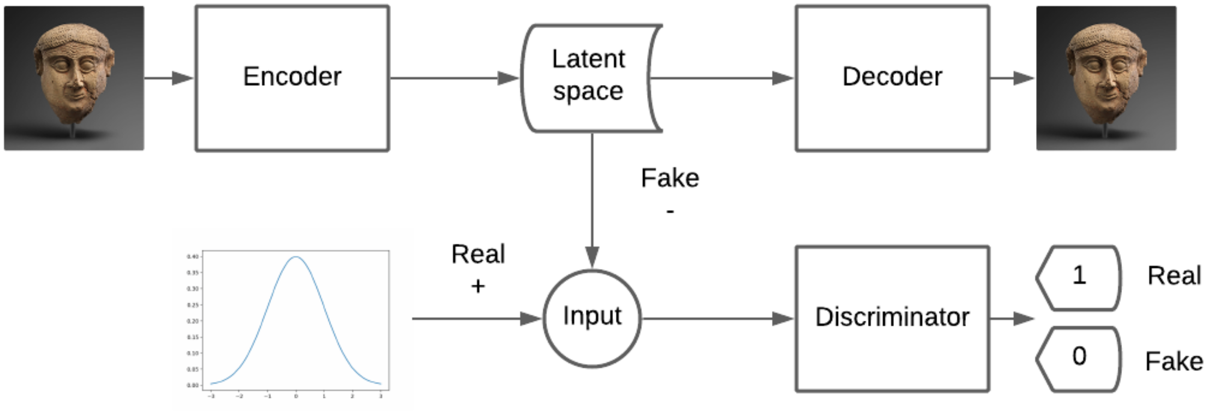

Fig. 1. Adverserial autoencoder model.

matching the hidden code vector of the autoencoder with a prior distribution. This can be used to map the input images into a model that follows a specified distribution, yielding a latent space that can be visualized in a 2D graph. The autoencoder is trained with a dual objective, one is the reconstruction as a traditional autoencoder, and the other is the adverserial training that matches the distribution of the latent representation of the autoencoder to an arbitrary prior distribution. The encoder learns to convert the data distribution to the prior distribution, and the decoder learns a deep generative model that maps the prior distribution to the data distribution. The adverserial autoencoders are useful in applications such as semi-supervised classification, disentangling style and content of images, unsupervised clustering, dimensionality reduction and data visualization. [4]

The visualization of the data is one of the important features of adversarial autoencoders, this characteristic can help to find visual tendencies of a particular dataset. This is useful in this field because historical art researchers can also aid their analysis with this visual representation of art to analyze influences or similarities found by the model.

### B. Structure

The model is very similiar to a simple autoencoder, but the encoder is trained in an adversarial way to force it to output a required normal distribution. The model is composed of 3 networks, the autoencoder (encoder, decoder), the discriminator, and the generator (encoder). The model is trained in two different phases: the reconstruction phase, and the regularization phase. In the former, the encoder and decoder are trained to minimize the reconstruction loss, which is the mean squared error between the input and the decoder output image. Therefore, backpropagation is performed through the encoder and decoder weights to reduce the loss. In this phase the discriminator remains unaffected. In the latter, the discriminator and the generator (encoder) are trained. The discriminator learns to classify the encoder output (negative samples) and the input from the real normal distribution (positive samples). The final step in this phase is to train the generator to output the prior distribution. To achieve this, the encoder output is connected as the input of the discriminator, while the discriminator target is set to positive samples and their weights become fixed. In this way, backpropagation is performed only through the encoder weights, which causes the encoder to reproduce the required distribution by looking to the discriminator weights. [5]

## III. Implementation

### A. Dataset

The dataset was provided by the Department of Art History of SMU, Dallas. The dataset included 4,315 images of figurines from different cultures in *.jpg and *.TIF format. It contained 3,144 files from the Hellenistic Babylonian culture, 887 files from the Classical Greek and Hellenistic culture, 204 from the Neo-Babylonian culture, and 80 files from the Achaemenid culture. The Hellenistic Babylonian dataset is further divided into 8 locations (Babylon, Seleucia of the Tigris, Uruk, Nippur, Ctesiphon, Kish, Borsippa, and an unprovinanced zone). From the PyTorch library, the DataLoader was used to import the images in batches of 100 images. To help the training of the model, the data was transformed to have a mean and standard deviation of 0.5 on the three channels of the image.

### B. Procedure and Algorithm

For this project two libraries were used to code the adverserial autoencoder model: tensorflow 2 and pytorch. The two libraries were used independently in two different codes and both were tested separately to look which one yield the best results. Pytorch showed a better performance compared to tensorflow. Although the tensorflow implementation was running correctly, the generator seemed to collapse very quickly into a single point of the distribution.Therefore, the procedure will be focused on the pytorch implementation. The algorithm

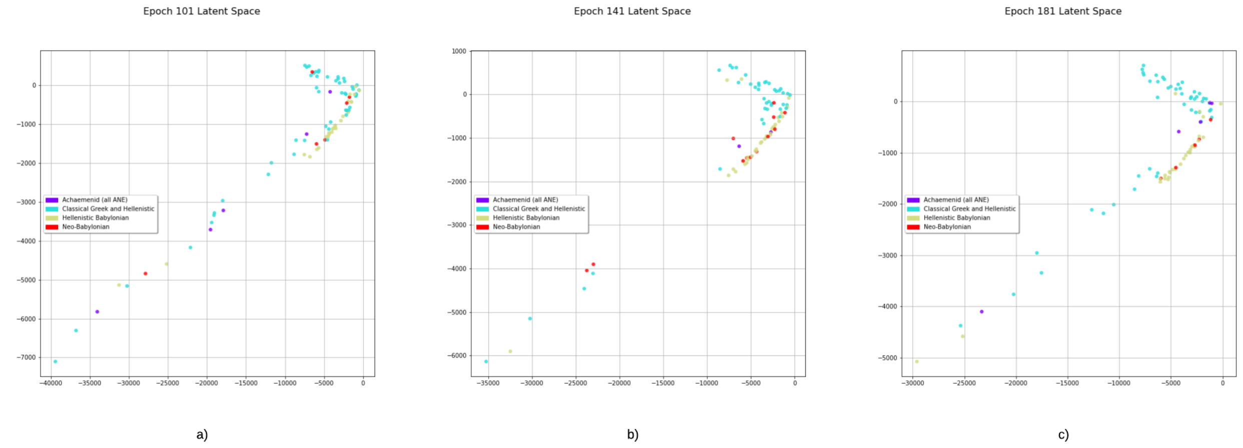

Fig. 2. Latent Space distribution throughout three epochs.

was made according to the model mentioned in the structure section and considering previous research and codes found in the literature. The algorithm was divided according to the next sections:

*1) Base parameters definition:* The base parameters were used during the whole procedure, which define the characteristics for the training of the network.The next list shows the parameters:

1) batch size = 100
2) height = 256
3) width = 256
4) channels = 3
5) input dimension = $height * width * channels$
6) number of units in dense layers = 1000
7) latent space dimensions = 2
8) epochs = 200
9) learning rate = 0.001
10) beta = 0.9
11) optimizer = Adam
12) prior distribution = normal (gaussian) $\mu = 0.0$ and $\sigma = 5.0$

*2) Network definition:* The next step was to create the networks for the project. The encoder and decoder pair were created very similar to each other. The encoder received an input with size equal to the input dimension, then the input was processed with two dense layers, both with 1000 units. The encoder output was a 2D tensor containing the important features that create the latent representation. The decoder takes the output from the encoder as the input to the network and makes the same process of the encoder in the opposite order. The output from the decoder has the same shape and size as the input tensor, which usually could be reshaped to create an image. This two small networks conform the whole network called autoencoder.

The discriminator is also another network consisting on 2 dense layers with 1000 units, both layers use relu as activation. The last and third layer is also a dense layer, but only with 1 unit, which create the predictions for real samples 1 or fake samples 0.

The generator is nothing but the encoder itself. The generator is treated as another network, because is part of the generative model. This will be trained in an adversarial manner with the discriminator to force it to output the prior distribution.

*3) Loss definition:* The losses for each network were defined as in the model. The autoencoder uses a mean squared error to minimize the reconstruction loss between the input and output of the decoder. The discriminator loss is the sum of the cross entropy loss of the real distribution and the cross entropy loss of the fake distribution. Finally, the generator loss is the cross entropy loss of the encoder output passed through the discriminator and the target set to 1, when the discriminator weights are fixed.

*4) Training:* The network was trained for 200 epochs, using 4135 images in batches of 100 images. As mentioned above the autoencoder was trained first, then the discriminator with the encoder, and finally the generator. The training last about 10 hours. During each epoch, the autoencoder loss, discriminator loss, and the generator loss was collected. Moreover, every 20 epochs a plot was made, showing the partial results of the latent representation.

## IV. Results

### A. The algorithm

The dataset was encoded after the training of the algorithm. The results were graphed to show the latent space of the images in Figure 3. Each of the civilizations is represented by a color, indicating that a figurine image was encoded to that point in the latent space. The latent space is not following the expected distribution described in the model, clustering at really high levels and producing points in the latent space far away from the [-5,5] range. This leads to the belief that the encoder is not trained enough and has not yet learned the

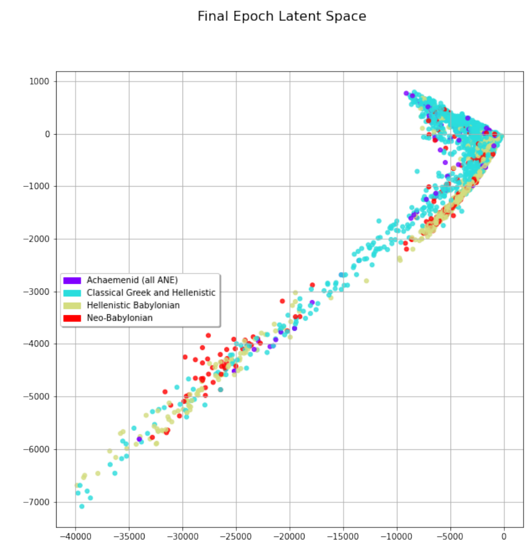

Fig. 3. Dataset Encoded into Latent Space

desired distribution. This can also be supported by looking at the losses throughout the training on Figure 4. Throughout the training, the loss values for discriminator was constantly high, while that for the generator was constantly low. The discriminator was not trained enough either, not being able to predict from a real distribution and the encoder's output. Because of this, the generator had no problem making a latent space that could trick the discriminator, making the generated distribution very different from the desired one. The autoencoder overall did not perform very well.

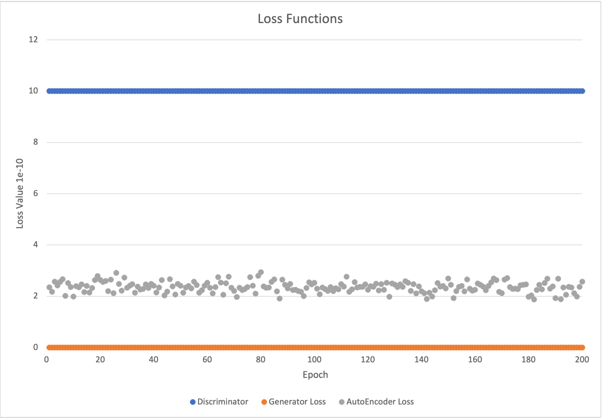

Fig. 4. Loss Values Throughout the Epochs

### B. Interpretation

The latent space that was shown in Figure 3 does not cover the whole latent space and the autoencoder had poor performance. But even with these results, a subjective interpretation of the data will be made. There appears to be a cluster of points from images from Hellenistic Babylonian in the ranges of [-1000, 0] in the x axis that differ from the cluster of points from images from the Classical Greek and Hellenistic in this same range. This tendency could be seen throughout the training, as shown in Figure 2, which shows the graphs from three different epochs of the encoded images from the final batch in the latent space. With more training, and a model that has results with better loss values, the Hellenistic Babylonian and the Classical Greek and Hellenistic figurines could be graphed in a separate are in the latent space, making art from these periods and cultures distinguishable from one another by the algorithm.

It is also notable that the Neo-Babylonian image points seem to cluster around points near the Hellenistic Babylonian points. Therefore, throughout the range of [-2000,-1000] in the x axis of the latent space only points from the encoded Classical Greek and Hellenistic images were produced. This could be an early indication that Hellenistic Babylonian figurine styles were influenced by Neo-Babylonian figurine styles. In the same way, the Achaemenid latent space points do not appear to overlap or cluster in the same area as the Hellenistic Babylonian, indicating that there was little cultural influence in the Hellenistic Babylonian art from them.

These interpretations are based on an early model of the AAE. Future work is needed to make a model that generates the latent space points in the desired distribution and that has a lower autoencoder loss function value. With this, any interpretation made from these results could be supported.

## V. Future Work

Although the model lacked training, the interpretations made from the results encourage the further study of this topic to train a model that can help art and historic researchers. Aside from the need for more epochs, the model also needs techniques that can help the discriminator to train better, this could also help the generator try to generate distribution that are more similar to the desired distribution. Another improvement would also be a more balanced dataset. More than 70 percent of the images from the dataset belonged to the Hellenistic Babylonian, while the images that belonged to the Achaemenid were a little below 2 percent.

## VI. Conclusion

The use of AAE can help researchers in the historic art field understand better the influences past civilizations had on one another. Although the model used in this paper did not yield the desired results, it does motivate further work to be done with improvements in the training model as well as the dataset to be used. Although the interpretation of said results would be analyzed by researchers, algorithms can help accelerate the understanding of their work and to find automatic patterns in figurines which can be used to characterize art and determine the relationship between artistic representations in each past civilization.

# References

[1] S. M. Langin-Hooper, *Introduction*. Cambridge University Press, 2020, p. 1–12.

[2] H. D. Baker, "The neo-babylonian empire," *A Companion to the Archaeology of the Ancient Near East*, vol. 1, pp. 914–931, 2012.

[3] B. N. Porter *et al.*, *Images, power, and politics: Figurative aspects of Esarhaddon's Babylonian policy*. American Philosophical Society, 1993, vol. 208.

[4] A. Makhzani, J. Shlens, N. Jaitly, I. Goodfellow, and B. Frey, "Adversarial autoencoders," *arXiv preprint arXiv:1511.05644*, 2015.

[5] N. Nagabushan, "A wizard's guide to adversarial autoencoders: Part 2, exploring latent space with adversarial..." Nov 2017. [Online]. Available: https://towardsdatascience.com/a-wizards-guide-to-adversarial-autoencoders-part-2-exploring-latent-space-with-adversarial-2d53a6f8a4f9