HIGH SCHOOL EDITION
Journal of Student Research

Volume 11 Issue 1 (2022)

AI-Powered Archaeology: Determining the Origin Culture of Various Ancient Artifacts Using Machine Learning

Archith Iyer[1] and Dr. Manoj Franklin[2#]

[1]Mission San Jose High School
[2]University of Maryland College Park
[#]Advisor

ABSTRACT

This paper attempts to determine the degree of accuracy to which we can originate and classify ancient artifacts using AI. It compares the performance of two CNN models (InceptionResNetV2, and a VGG-19) on a specially curated dataset of 55,000 ancient artifact images belonging to over 343 different cultures. The dataset consists of samples from different regions all over the world and is built using the British Museum's online collection of over 3 million artifact images. For training the CNNs, an NVIDIA RTX A6000 GPU was used on a cloud computing platform, achieving high accuracy for both models. Two trials were run for the InceptionResNetV2 model, which achieved a final validation accuracy of 65 percent for its best trial. The VGG-19 performed poorly on the dataset, only achieving a 12.95 percent validation accuracy. Afterwards, the models obtained from both InceptionResNetV2 trials were evaluated on a set of artifacts of unknown origin, coming up with surprising results and providing insight into how effective image recognition can be in the field of archaeology.

Introduction

The time period that defines "Ancient History" generally expands from the first instances of cuneiform writing (3200 BCE) to the fall of the Roman Empire (476 AD). During this time period, civilizations developed in all regions of the world, leading to vastly different cultures. An ancient artifact is defined as any object made by the humans who lived during that period, while its origin refers to the civilization/culture that its creator(s) belonged to. Archaeologists have used various methods to determine the origin of ancient artifacts found during excavations of ancient sites. Methods such as carbon dating, material analysis, and analysis of artistic techniques used in its creation have proved wildly successful. These methods often work, because artifacts of the same culture hold many commonalities; for example, Roman busts are often white due to their use of marble in the creation of sculptures. Despite new advancements in the field of archaeology, one technological method remains uninvestigated: image recognition.

Image recognition has been revolutionized by the Convolutional Neural Network (CNN). CNNs utilize kernels, which slide over an image in order to extract features. These "convolutional layers' ' enable an image to maintain its 2D shape. After multiple convolutional layers, the remaining data is flattened and passed through a multilayer perceptron (dense layer) in order to finally classify the image. Currently, there exist many CNN architectures, each boasting high accuracy and benefits.

Using a dataset of ancient artifact images, separated into classes based on origin culture, it is possible to train a CNN model to accurately classify these artifacts provided their images.

Methods

The Dataset

All images in the training set were downloaded in batches from the British Museum's online collection, based on their region. At first, a total of 93,000 images were downloaded. The distribution of images based on region is summarized in Figure 1.

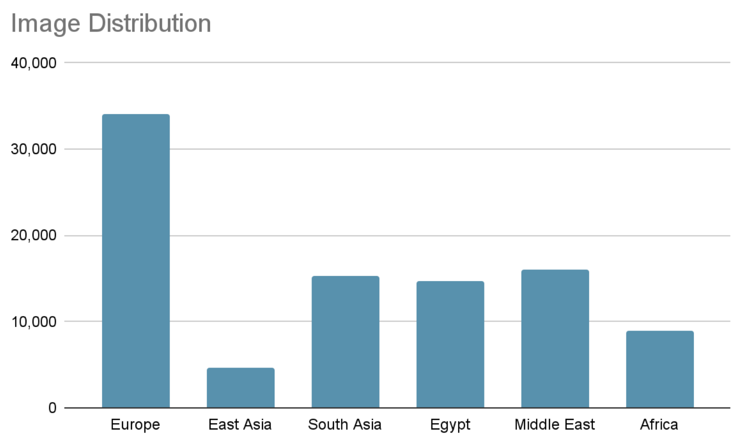

Figure 1. Distribution of images based on region

Due to the museum collection's lack of ancient artifacts from East Asia and Africa, a limited number of images from these regions were available for download. The 34,000 images downloaded for the Europe region is a sample from the available 200,000 images. This sampling was done to eliminate bias and ensure that the dataset would not contain too many images from Europe in comparison to other parts of the world. Sample images from the dataset can be observed in Figure 2.

Figure 2. From left to right: Ancient Roman Coins, Greek Amphora, Zhou Dynasty Jade Pieces

All images of artifacts were grouped into folders based on their respective culture. In order to eliminate bad data and improve model accuracy, all classes containing less than five images were thrown out. Afterwards, all corrupted images were deleted.

After processing, only 55,000 images, belonging to 343 different classes, remained in the dataset. 44,000 were used in the training process, while the other 11,000 were used for validation, or to test the performance of the model.

Before being fed into the CNN models for training, the images were zoomed in by a factor of 20 percent in order for the model to focus on the artifact instead of the white borders surrounding the image. Afterwards, for augmentation, a horizontal flip and a shear were performed. The horizontal flip adds to the dataset, by reflecting the image along the X-axis in order to mimic how humans observe it at different angles. The shearing further adds to the dataset by distorting the image along the X-axis. The zoom and horizontal flip effect can be observed in Figure 3.

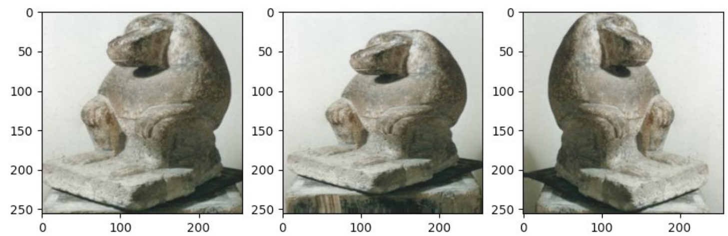

Figure 3. Zoom and horizontal flip applied to an Egyptian Sphinx

The Networks

Network 1: The InceptionResNetV2

The architecture of the InceptionResNetV2 network can be summarized in figure 4.

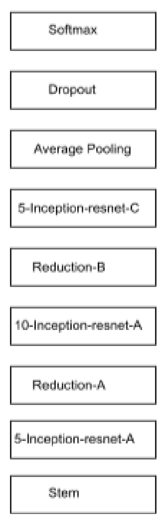

Figure 4. Structure of the InceptionResNetV2

The InceptionResNetV2 takes in images of size 299 by 299. The network consists of inception blocks: An image model block that allows us to use many different filter sizes(ex. 5 by 5, 3 by 3, etc.) and pooling operations in a single block. The inception blocks also utilize 1 by 1 convolution, which allow the image to maintain its dimension while passing through layers. The Inception blocks are connected by residual connections, which allow images to pass through the network without the use of activation functions. At the end of the network, 20 percent of the nodes are dropped out (to provide regularization and prevent an overfit), before being passed through the softmax function to yield the class probabilities.

Network 2: The VGG-19

The purpose of the VGG-19 was to act as a control, to see if decreasing the sophistication of the network would yield significantly worse results, as the VGG-19 was repeatedly outperformed by newer Inception and Resnet models. The architecture of the VGG-19 network can be summarized in figure 5.

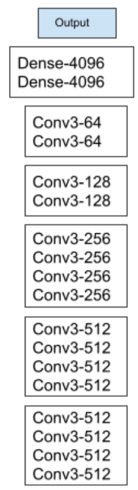

Figure 5. Architecture of the VGG-19

The VGG-19 has a very basic architecture. It takes in images of size 224 by 224, and utilizes 3 by 3 convolutions. The number of filters starts off at 64 for two convolutional layers, and then increases to 128 following a max pooling operation. The process repeats, but after doubling again to 256, the number of layers per filter-count increases to 4. From here, convolutional layers with 512 filters are added, followed by a max pooling operation, and then another set of 512-filter convolutional layers. At the end, the processed image is flattened, before passing through 2 sets of 4096 node dense layers, and then the output layer.

The Trials and Training

Two trials were run for the trials involving the InceptionResNetV2 while only one was run for the original model. The second trial for the InceptionResNetV2 served to improve validation accuracy by fixing errors(such as an overfit) from the first trial.

Initially, 44,000 training images were evaluated against 11,000 images in the validation set. For the first trial, the initial batch size was set to 32, the “Adam” optimizer was used, and shuffling was not enabled for the training set.

All models were trained and evaluated on a NVIDIA RTX A6000 on a cloud computing platform.

Results and Discussion

. The performance of the CNN’s on the dataset, along with the possible evaluation of artifacts of unknown origin are discussed in this section.

Network 1: The InceptionResNetV2

Trial 1

The training of the network can be summarized in figure 6 below.

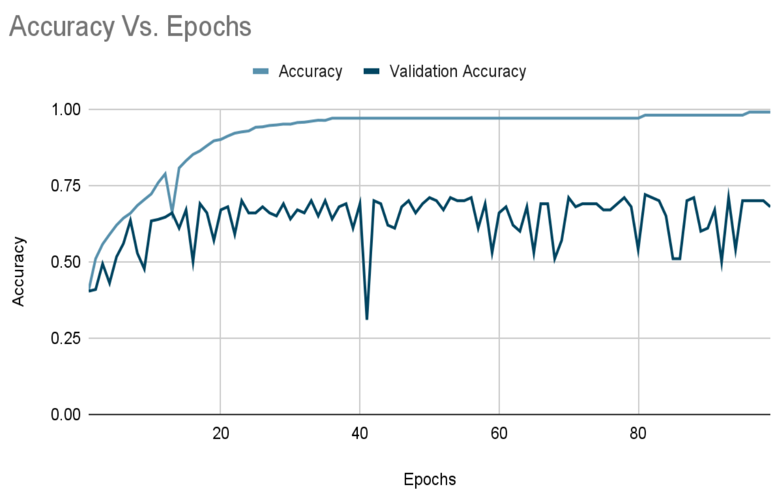

Figure 6. Accuracy vs Epochs for InceptionResNetV2(trial 1)

As observed in the above graph, the first trial yielded a severe overfit. While it initially increased, the validation accuracy "zig-zagged" between 60 and 70 percent starting around the 16th epoch and proceeding for the rest of the training. As the epochs increased, the "zig-zag" started to settle around the 70 percent mark, before dropping back down to yield a final 62.49 percent validation accuracy, as opposed to a 99 percent training accuracy. The overfit was corrected by shuffling the training set(to prevent the model from "memorizing the training set"), while also reducing the number of epochs from 100 to 60. Finally, l2 regularization was added to each layer of the network to penalize an overfit, therefore improving the validation accuracy of the model as opposed to training accuracy.

Trial 2

Trial 2 re-trains the InceptionResNetV2 but incorporates the changes mentioned above.

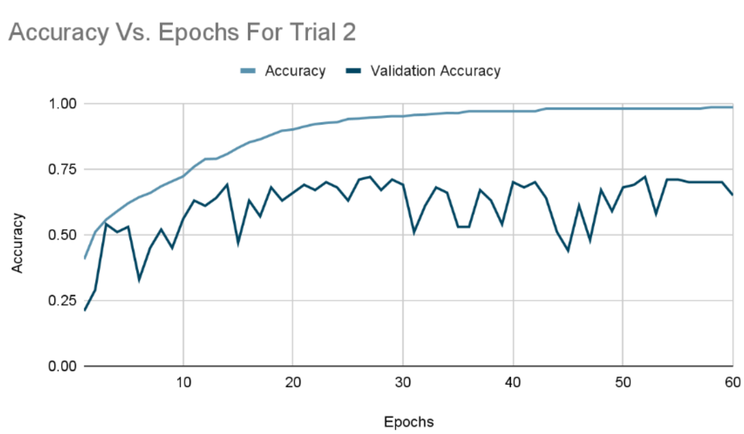

Figure 7. Accuracy vs Epochs for InceptionResNetV2 (trial 2)

The fluctuation in validation accuracy is slightly less for this trial, as there are extended periods of consistently followed by sudden decreases for a single epoch. The final validation accuracy was 65 percent, a 3 percent increase from the first trial.

Network 2: The VGG-19 Model

The training of the VGG-19 can be summarized below.

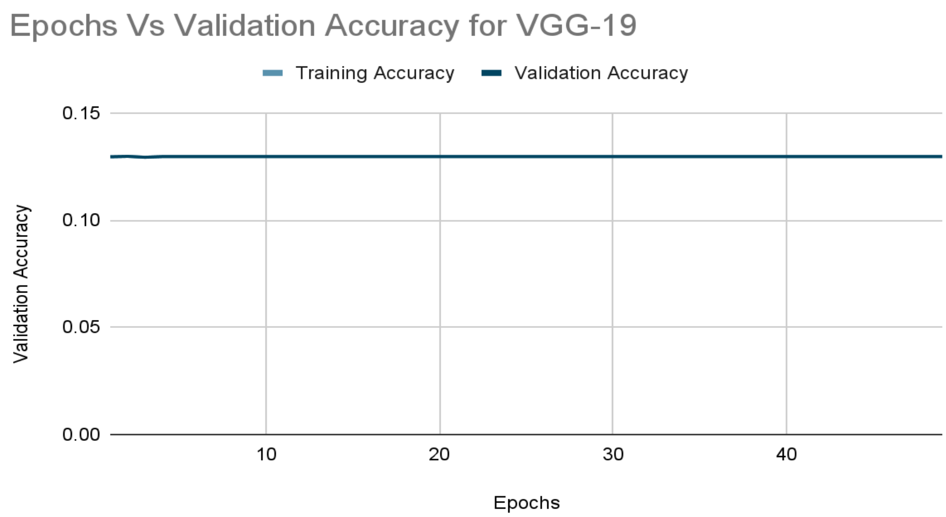

Figure 8. Accuracy vs Epochs for the VGG-19 Model

Both the training and validation accuracy fluctuated around 12.9 percent, increasing and decreasing by approximately 0.01 to 0.05 percent per epoch. The reason for this low accuracy, as well as the lack of learning, is unknown. Changing the learning rate from 0.001 to 0.01 and again from 0.01 to 0.0001 yielded similar results. Perhaps the VGG-19 Model is unsuited for this classification task, and hopefully future research will resolve this issue.

Artifacts of Unknown Origin

The InceptionResNetV2 models were then evaluated on a set of artifacts of unknown or questionable origin. Each of the models' top three choices were obtained by running the models on the image, and then displaying the classes with the three highest probabilities. In the table below, the choices are ranked from top choice to third best choice.

Evaluation of Artifacts of Unknown Origin

Image of Artifact	InceptionResNetV2 Trial 1	InceptionResNetV2 Trial2
	1. Roman Imperial 2. East Greek 3. Pre-Imperial Rome	1. Roman Imperial 2. Late Helladic(Early Greek) 3. Indus Valley Civilization
	1. Ancient Greek 2. Hellenistic 3. Mesopotamian	1. Romano-British 2. Imperial Roman 3. Pre-Imperial Roman
	1. Akkadian 2. Roman 3. Classical Greek	1. Neo Assyrian 2. Old Babylonian 3. Late Babylonian

The Top Image

The top image appears to be a coin, bearing the face of an important figure. Such designs were common among the Romans during the Imperial period, therefore both trials' first choice(Roman Imperial) is most likely the true origin culture of this artifact. Upon analyzing the class probability distribution of the first trial, the probability that the artifact was Imperial Roman was relatively high (48 percent). Since the model was so confident in its first choice, we can ignore its second and third picks.

The Middle Image

This artifact was found during an excavation of a third-century BC temple near the ancient city of Troy. Since there currently exists no "trojan" culture in the British Museum's database, the first trial's top two choices were its nearest relatives, Ancient Greek and Hellenistic cultures, which flourished in the regions surrounding the city of Troy during that time period. The second Inception trial's prediction that the artifact is Roman, is slightly less plausible. The image depicted in the artifact is that of the goddess Athena, one not worshipped by the Ancient Romans. Furthermore, the artifact was found in a region (Turkey) significantly more influenced by Greek culture than Roman culture.

The Bottom Image

This artifact remains a tricky one to classify. The model's top choice of Akkadian is certainly plausible, as it bears similarities to Akkadian sculptures (for example, as seen in figure 9 below). The other two choices for the trial one model though, are not reasonable, as this figure bears little resemblance to either Greek or Roman busts. The second trial yielded slightly more reasonable predictions. The Babylonian and Assyrian cultures also produced sculptures which resemble this artifact. Furthermore, as opposed to only the first of the top three choices being reasonable, the second trial resulted in all three of the top three choices being possible origin cultures for this artifact. One thing both models agree on is that the artifact is likely from an ancient Middle Eastern culture, as the Akkadians, Babylonians, and Assyrians resided in the modern-day countries of Iraq, Iran, and Turkey.

Figure 9. Example of Assyrian sculpture(right) and Akkadian sculpture(left)

Conclusion

As the field of Artificial Intelligence evolves, new purposes are being found for it. In this case, AI is successfully used to approximate the origin culture of ancient artifacts obtained from excavation sites. With images of artifacts widely available online, such as on the British Museum's website. a dataset of these images grouped into classes took little difficulty to create. While basic CNN models are not sophisticated to be trained on this dataset, recent developments in CNNs enable us to use more powerful models, such as InceptionResNetV2, to obtain significantly high training and validation accuracy. While regularization techniques improved performance of the model on the validation set, they provided similar insights on artifacts of unknown origin as the non-regularized model. The top-1 choices for both models on the artifacts of unknown origin were not only reasonable, but likely the true origin culture of that artifact.

References

5 techniques to prevent overfitting in neural networks. KDnuggets. (n.d.). Retrieved December 1, 2021, from https://www.kdnuggets.com/2019/12/5-techniques-prevent-overfitting-neural-networks.html.

Allibhai, E. (2018, November 21). *Building a deep learning model using Keras*. Medium. Retrieved December 1, 2021, from https://towardsdatascience.com/building-a-deep-learning-model-using-keras-1548ca149d37.

Canuma, P. (2018, October 11). *Towards Data Science*. Retrieved from Image Pre-processing: https://towardsdatascience.com/image-pre-processing-c1aec0be3edf

Collection. The British Museum. (n.d.). Retrieved December 1, 2021, from https://www.britishmuseum.org/collection.

Elhamraoui, Z. (2020, May 16). *INCEPTIONRESNETV2 simple introduction*. Medium. Retrieved December 1, 2021, from https://medium.com/@zahraelhamraoui1997/inceptionresnetv2-simple-introduction-9a2000edcdb6.

How to create a graph in a google doc. Alphr. (n.d.). Retrieved December 1, 2021, from https://www.alphr.com/how-to-create-graph-google-doc/.

The search for the Lost City of troy - british museum blog. British Museum Blog - Explore stories from the Museum. (2021, August 24). Retrieved December 1, 2021, from https://blog.britishmuseum.org/the-search-for-the-lost-city-of-troy/.

Team, K. (n.d.). *Simple. flexible. powerful.* Keras. Retrieved December 1, 2021, from https://keras.io/.

Wikimedia Foundation. (2021, November 30). *Ancient history*. Wikipedia. Retrieved December 1, 2021, from https://en.wikipedia.org/wiki/Ancient_history.