




“I was painted by...”: A Case Study on the Use of CNNs for Image Classification in the Humanities

Marta Kipke <marta_dot_kipke_at_uni-goettingen_dot_de>, Institut für Digital Humanities, Georg-August-Universität Göttingen  <https://orcid.org/0000-0003-0130-261X>

Lukas Brinkmeyer <lukas_dot_brinkmeyer_at_isml_dot_de>, Information Systems and Machine Learning Lab, Stiftung Universität Hildesheim  <https://orcid.org/0000-0001-5754-1746>

Martin Langer <martin_dot_langner_at_uni-goettingen_dot_de>, Institut für Digital Humanities, Georg-August-Universität Göttingen  <https://orcid.org/0000-0002-6109-4878>

Lars Schmidt-Thieme <schmidt-thieme_at_isml_dot_uni-hildesheim_dot_de>, Information Systems and Machine Learning Lab, Stiftung Universität Hildesheim  <https://orcid.org/0000-0001-5729-6023>

DOI: <https://doi.org/10.63744/a8rvhq9zmk>

Abstract

EGRAPHSEN is a case study on image classification in the humanities, specifically on painter attribution on Attic vase paintings. This study aimed to explore the new perspective that artificial intelligence (AI) can offer when studying traditional methods and heterogeneous domains. When we translate the task (painter attribution), we have to consider the idiosyncrasies of the data domain (Attic vase paintings). This is challenging for both, classical archaeologists and computer scientists. In this paper, we address how to approach the challenges in the creation of the dataset. We carefully selected and prepared the data, reflected on potential biases and trained a convolutional neural network (CNN) accordingly. Specifically, we developed sampling criteria to combat the biases and a hierarchical labelling system to segment the images into details. Our model architecture was designed to process sets of images instead of only one individual image, which enables us to experiment with different combinations of image segments. This forms the basis for an analysis framework, which allows us to go beyond mere painter attribution and to explore the ambiguity of image similarity itself.

1. Introduction

When visitors look at a painting in a museum, they might expect a clear answer to the question “who painted it?”. Often a painting contains a signature, which makes the answer to this question seem pretty obvious. However, the answer is not always straightforward. Sometimes the painter cannot be pinpointed to one specific individual, and our knowledge might be limited to a timeframe, a region, a workshop. Even well-known paintings were not always signed. A good example is the *Sleeping Venus*. It was first thought to be from the school of Titian and was later identified to be by Giorgione [Anderson 1996]; [Lüdemann 2005]; [Uglow 2014]. Identifying a painter is a riddle, waiting to be solved by Art Historians and Classical Archaeologists. In our project EGRAPHSEN ^[1] we view this question from the perspective of artificial intelligence (AI): Can AI help solving this riddle? What does this reveal about the task itself? And more importantly: How can artificial intelligence help us go beyond that and answer fundamental questions regarding image similarity and variation? How exactly does an artificial neural network differ from a human scholar? Painter attribution – the task of identifying painters on unsigned images – is a complex field with many differing opinions, which poses a challenge to the operationalisation of the task from the data selection up to the interpretation of the results.

The overall situation is complicated: It was not always common practice to sign paintings. Even a famous painter may have left some paintings unsigned, for reasons which we can only speculate about [Hegener and Horsthemke 2013]; [Burg 2007]. Nor is an existing signature necessarily a confirmation of authorship, as it may have been forged for profit [Jones and Craddock 1990]. Due to atelier workflows we cannot take real historical signatures at face value either: Was it the actual painter who signed a work, or the leader of the workshop [Liedtke 2004], [Büttner 2017]? Overall, painter attribution can be understood as a tool which paves the way for further research on historical contexts, workshop processes and style in general.

In EGRAPHSEN we approach this task at the very core: Our main focus lies in understanding painter attribution as a method, especially its manifestation in Classical archaeology. Our use case are Attic vase paintings from Greek antiquity, which pose an even greater challenge than later canvas paintings, as the former are only rarely signed at all. With this, other hints to solving this riddle become increasingly important, which in this case are primarily found in the visual comparison of stylistic phenomena. By studying the artists styles a scholar can “develop an eye” for the respective artists. As fundamental as this method is, this skill is subjective and not always verifiable. The expert’s eye has to be trained and for an untrained eye the arguments might seem elusive and presumptuous, begging the question if they hold any value at all. Throughout the times the value of connoisseurship has been questioned in both Art History and Classical Archaeology (for the classical archaeological perspective see: [Graepler 2016]; [Sparkes 2013, 90–113]. The discussion is extensive and while painter attribution has its defenders [Oakley 1998], it is still controversial and challenging to navigate. Despite of this, the method is still practiced in major publications recently. ^[2] The insight it offers is too valuable, making it a good use case to revise and discuss traditional methods through the lens of a novel computational approach.

With this, we are also researching the role of AI in humanities disciplines, including its metrics and its dangers. Humans are complex, as is humanities research and the material we study. Can a computer *understand* humans or art? This is a controversial question, which leads to heated discussions between computer scientists and art historians [Wasielewski 2023]. Our work adds a case study which emphasizes the complexity of humanities research questions and the necessity of an informed and reflective approach, shaping the data compilation closely to our research questions.

Our use case is particularly suitable because of its dual nature: On the one hand it is strictly about shapes and lines: the style of a painting. We call them **formal characteristics**, with “form” meaning the lines and shapes, the design, arrangement and style. This describes a phenomenological approach to images that should not be confused with formalism as a scholarly direction of the late 20th century [Pinotti 2012]. These characteristics ultimately help answer the question: “**How** is a painting painted?” in contrast to **what** is painted. It seems that neither a human nor a computer needs additional information beyond the painting itself to be able to “learn” connoisseurship. On the other hand, painter attribution includes subjectivity and visual association and therefore uncertainty, fuzziness and ambiguity. These concepts are difficult to grasp with computational methods, but they are essential for the method of painter attribution as well as many research

1

2

3

4

5

questions in the humanities in general. At first glance, it seems that painter attribution could be easily translated into a machine learning classification set-up, with the different painters representing the classes. However, human connoisseurship is inherently instinctive and algorithms require specificity. A neural network needs distinct and well-defined categories to function, but the data is subject to biases and distortion. In the end, additional information beyond the painting might prove more important than one might anticipate.

This research promises new insights on both sides: As Classical Archaeologists we re-evaluate our view on painter attribution and the underlying concept of image similarity through the lens of AI, while Computer Scientists re-evaluate their algorithms using our complex and heterogenic data. Thus, our research goal is as follows: We aim to evaluate and develop methods for image-related research questions with the help of artificial intelligence. We do not aim at recreating painter attribution through computational means; instead we re-evaluate it critically and beyond its surface. In EGRAPHSEN we developed methods to (1) compile our data in a meaningful way and (2) explore image similarity with computational means. In this paper we are going to present the first part of our research and provide a preliminary insight into our results.

Working with AI presents major challenges when it comes to the translation of humanities knowledge into data – and conversely, it is also difficult to generate knowledge from experiment results. For the purpose of this paper, our research question can be divided into the following sub-questions:

- How do we operationalize the traditional method painter attribution and how does it relate to the algorithms used?
- What biases must be taken into account when compiling the dataset, and to what extent can we control the input in order to find an adequate data representation for the purpose of attribution?

Overall, these questions set the baseline for the interpretation and exploration of the model in regards to uncertainty and fuzziness of visual references and relations.

In a first step we need to contextualize our research questions, both for the traditional method (2.1) as well as in regards to other computational painter attribution projects (2.2) and machine learning in general (2.3). Doing so, we want to demonstrate the methodological demands of the task. The second step concerns the sampling strategies on different levels of input – the selection of painters (3) and the selection of input segments (4). After a short interim conclusion (5), we present our experimental set-up and model architecture (6), where we outline the main challenges of the algorithms. We then provide a first analysis of our results, which outlines our model's performance and first insights into the similarity networks between the painters (7). A final conclusion explores possibilities for further research, including an outlook on the analysis part of our work (8).

2. Painter Attribution in Context

2.1 Beazley and Painter Attribution in Classical Archaeology

In order better to understand the skill of connoisseurship and the visual relations between the paintings and their styles, we first need to outline the traditional method. How do Classical Archaeologists attribute painters on Attic vases?

We study an era so long ago that some aspects of the culture have been distorted and lost. Yet others have been carefully preserved. The ancient world has fascinated many people throughout the times, sometimes even taking on an exemplary character. This includes all kinds of cultural property: history, philosophy and literature. As Classical Archaeologists we study the appearance, execution, creation and contexts of *material culture*, developing methods to describe the developments of imagery, objects and style. For this purpose, we have to correlate them with each other, which results in intricate chronological and typological systems (on the dating of vases specifically see [Schmidt (ed.) 2022]). One interest of painter attribution lies in the creation of such an overarching system based on individual style, but also covering chronological aspects.

The interest in the objects themselves is manifold due to the insight they give us into the realities and ideas of life in ancient Greek society. One aspect is their use in everyday life, for example as grave goods [Tsingarida (ed.) 2009]. Another aspect concerns the imagery they depict. The themes range from mythological stories to portrayals of everyday Athenian citizens [Bérard and Vernant 1985]. This gives us an insight into the values, ideas and culture of that time. Research on Attic vase paintings is vast, as witnessed by numerous introductory handbooks on the subject [Vierneisel (ed.) 1990]; [Robertson 1992a]; [Scheibler 1995]; [Boardman 2001]; [Mannack 2012]; [Cook 2013 ed.]; [Sparkes 2013]. Another interest lies in artisan production and craftsmanship, which is closely tied to the question of painter attribution [Coulson et al. 1997]; [Palagia and Oakley (eds.) 1997]; [Oakley 2014]; [Eschbah and Schmidt 2016].

While some artists' names, for example those of sculptors, are known from historical sources [3], we do not know much about vase painters historically. Sometimes, however, there is a Greek name painted on a vase, accompanied by the phrase ΕΓΡΑΦΣΕΝ (I was painted by), meaning that the person with that name painted that particular vase [Osborne 2010]; [Pevnick 2010]; [Hurwit 2015]. These signatures were tempting clues that fuelled scholars to study the painters of the vases as if they were great masters. The deep appreciation for the paintings and the challenge of analysing their stylistic details led to decades of research on painter attribution in 19th-20th century Classical Archaeology [Schmidt 2016]. The question "*Who painted this vase?*" became hard to resist.

One particular scholar, Sir John Davidson Beazley (1885-1970), emerged as an authority in this field. Beazley developed his skills at the height of painter attribution research during the first half of the 20th century [Rouet 2001]. But his method is often retraced to earlier studies of Renaissance paintings and the specific approach developed by the art historian and doctor of human anatomy Giovanni Morelli (1816-1891, see [Kurtz 1986]). Morelli theorized about painter attribution and developed a specific approach for this task. It was his approach, that led to the *Sleeping Venus* being recognized as a painting by Giorgione. Morelli prided himself on being an anatomist and his method consisted of a systematic and refined comparison of anatomical details between the alleged paintings of an artist. Specifically those details that might seem insignificant at first glance turn out to be particularly important, because this is where the individuality of the artists is most clearly expressed [Wollheim 1973]; [Ginzburg and Davin 1980]; [Vakkari 2001]; [Hinojosa 2009]; [Uglow 2014]. Regardless of whether Beazley actually followed Morelli specifically, it is conceivable that Attic vase painters could be distinguished by such characteristics as well. The craftsmen produced multiple paintings per day and were routinely handling their tools, making shapes flow easily from their hands.

By studying the signed works of a painter, scholars were able to grasp a painter's *handwriting*, with details like eyes, hands or ankles being the *letters*. Looking at a signed cup (Figure 1 [4]), we are able gain a first impression of the painter Douris. His signatures survived in relatively large numbers and as such his style is well studied and established [Buitron-Oliver 1995]. Looking at the pictures, we are often drawn to the faces first. It is possible to discern the similarity between eyes or mouths, for example. Their context is also important: How are the features arranged in the face? How big or small are they? Are they scrunched together or more spaced out? With this, we look beyond mere *letters*, we also consider the context; the words they are used in. Scholars have of course examined not only

faces, but other body parts as well. In fact, the monograph on Douris features a section on the renderings of larger body parts of naked male figures, such as the chest and the legs [Buitron-Oliver 1995].

Studying the paintings in great detail gives a good impression of a painter's idiosyncrasies. But it also reveals the **variation** of their 'handwriting'. Beazley was one of the first to shift the focus from signed works to unsigned works on a larger scale [Neer 1997]; [Sparkes 2013]; [Oakley 2009]; [Pérez 2018]; [Driscoll 2019]. Not only was he able to identify unsigned paintings from known painters, he even identified painters with no known signatures at all. He named them after a topic they painted (e. g. "painter of the Paris Gigantomachy") or the city of the museum with their most well-known work (e. g. "Berlin Painter"). This way, he created an extensive structure with hundreds of painters.

15

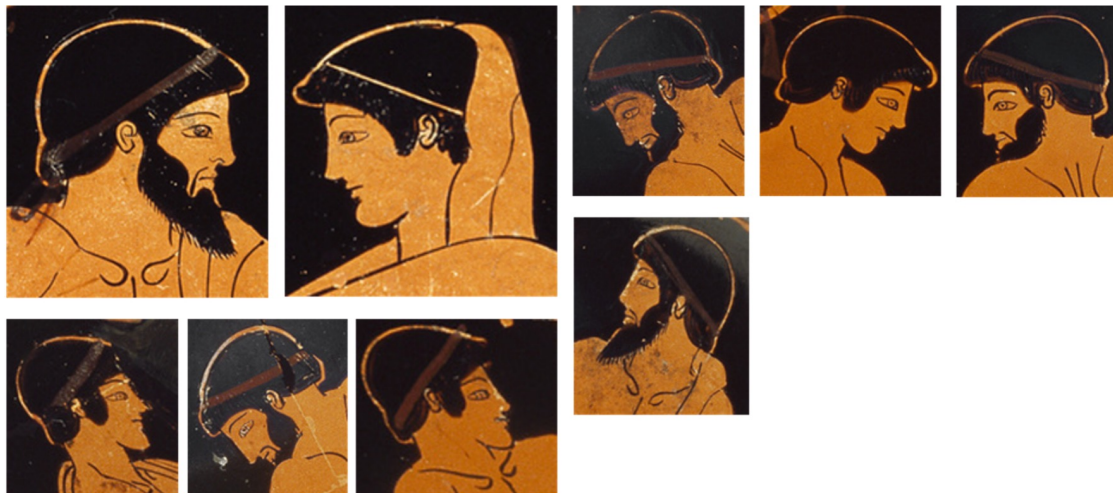


Figure 1. Faces from a signed cup by Douris. By studying the details and their variations scholars like Beazley are able to recognize the painters handwriting. BAPD 275972, Los Angeles Getty Museum of Art, 86.AE.29

To understand how he was able achieved that, we turn our attention to one of these painters. Figure 2 shows a few examples of the way the Berlin Painter drew faces. Beazley was especially fascinated by this painter, as attested by several publications [Beazley 1922][Beazley 1930]; [Beazley 1911]. In the first publication he describes the linework of a handful of vases in great detail. He not only looked at faces, but placed great importance on the rendering of the body and especially the drapery of the clothes. [5] A combination of all the features, including their deviations, led him to the conclusion that these particular vases must come from one hand. He writes "A **system** so definite, coherent, distinctive, and in some respects so wilful, is most easily intelligible as a personal system" and "the child, above all else, of one man's brain and will" [Beazley 1922]. He describes the painters' style as "**system of renderings**", which helps to understand his approach: Beazley compared not only stylistic details, but their overall interconnection, allowing him to not only identify known painters, but systems themselves. Beazley trained his eye well during his long-lasting career and sometimes he changed his mind about certain attributions. The more pictures he studied, the better his connoisseurship became.

16

Beazley's success was unparalleled and many of his observations, especially those regarding the Berlin Painter, still hold value today [Sparkes 2013]; [Oakley 2007]. His most important contributions remain the volumes of his Attic vase painter series, where he simply lists hundreds of thousands of vases, sorted by painter [Beazley 1963]; [Beazley 1956]. His approach itself became subject of study as well [Kurtz 1986]; [Neer 1997]. This demonstrates the ambivalent relationship of Classical Archaeology towards painter attribution.

17



Figure 2. Cut out faces from vases attributed to the Berlin Painter. Upper row: 201879B (New Haven, Yale University 1913.133), 201836A (London, British Museum 1843,1103.51, Image Credit: © The Trustees of the British Museum CC BY-NC-SA 4.0, Assetnumber: 1613148360), 201811B (New York, Metropolitan Museum 56.171.38), 275093 (New York, Metropolitan Museum 65.11.12) Lower row: 275093 (New York, Metropolitan Museum 65.11.12), 201836B (London, British Museum 1843,1103.51, Image Credit: © The Trustees of the British Museum CC BY-NC-SA 4.0, Assetnumber: 1613148365) 201829B (London, British Museum 1843,1103.74, Image Credit: © The Trustees of the British Museum CC BY-NC-SA 4.0, Assetnumber: 283160001), 202004 (New York, Metropolitan Museum 22.139.32)

While Beazley's contribution to Classical Archaeology is undeniable, he and especially his successors were harshly scrutinized, especially in the 1990s and early 2000s. Many aspects, but also the very existence of this approach were criticized for many reasons, some of them valid, others just coming from a different place of interest [Sparkes 2013]; [Driscoll 2019]; [Graepler 2016]; [Arrington 2017]. Some of the criticism is aimed at the language of painter attribution, which can come across as imprecise and judgemental. Beazley's early observations on the development of Douris for example reads as follows: "*Douris the youth [meaning in his earlier period] is a lively and graceful character: Douris the man is scrupulously neat, highly accomplished, sleek, decent and dull*" [Beazley 1918, 97]. Calling a painter dull might seem insulting, but for us, even such statements are useful, as we can still extract formal observations from them. The term "dull" might indicate a certain repetitive nature in his works, which is a valuable observation on the variation of the renderings.

18

Beazley's extensive work is fundamental for our own research, but not because we want to re-attribute works or re-analyse the works of certain painters. We are first and foremost interested in understanding these "systems of renderings" as a tool for categorizing visual similarity. At its core, Beazley's method is comparing one system of renderings against another, considering the variation within a painter's hand. In the end, it comes down to two questions: How much can a style vary to still belong to one painter's hand and how do we interpret specific kinds of similarity between paintings?

19

2.2 Painter Attribution: A Classification Task

Connoisseurship is often criticized for focusing too much on the shapes in an image, instead of its content. However, this is actually an advantage for machine learning purposes, since a strictly descriptive approach is something an algorithm can usually easily process. So of course, we are not the first to attempt and reflect painter attribution with computational methods. Most such attempts are rooted in Art Historical research of later time periods however, which come with different approaches to data and painting styles. Paintings in later periods are usually canvases covered in brushstrokes. Although vase painters used a brush as well, usually only the contours of the figures are painted and there are rarely large areas covered in a close pattern of brushstrokes. In this sense they are closer to drawings than post-antique paintings. Apart from this, the task of painter attribution is the same in its basic structure.^[6] Therefore, other projects might offer useful insights into sampling strategies and other challenges.

20

There have been large-scale approaches to train classifiers on individual or periodical style based on large datasets like WikiArt as early as 2011 (for a review see [Santos et al. 2021]). However, as Bell and Offert pointed out, the definition of style in such projects is not always in line with the definitions by Art Historians and Classical Archaeologists [Bell and Offert 2021]. This is a phenomenon apparent in many projects, but it becomes less prevalent the more precise the research questions are and the stronger they connect to traditional Art History research. Such in-depth projects often revolve around painters who either have controversial bodies of work or who are subjected to forgeries due to their popularity. While forgery detection is a common goal for many scholars in that field [Levy et al. 2014]; [Narvaez et al. 2022], there are also examples of more profound Art Historical research questions beyond that, for example on Leonardo da Vinci and the heavily restored Salvador Mundi ([Lewis 2019]; [Frank and Frank 2021]; [Langmead et al. 2021]; on the Salvador Mundi see: [Lewis 2019]) or Vincent van Gogh and his distinct brushstroke patterns ([Li et al. 2011]; [Folego et al. 2016]).

21

Despite varying degrees of success and methodological discussion, these approaches show that painter attribution translates especially well into supervised machine learning (SML) algorithms. It is a classification problem and SML algorithms treat given problems as a classification task. SML includes many methods, from statistical approaches to deep learning [Elmrabet et al. 2021]. For the purpose of image pattern recognition, deep convolutional neural networks (CNN) prove to be the most successful approach.

22

In SML, a model is trained on a set of labelled data to recognize patterns and make predictions on new, unseen data. In our case, the labelled data would be a set of images of ancient vases with known painter attribution, and the model would be trained to recognize the stylistic characteristics that distinguish each painter's work. Once the model is trained, it can be used to analyse new images of vases and predict the most likely painter based on the similarities with the previously labelled data. In order to ensure that the neural network is accurately identifying the work of individual painters and not just picking up on general trends, it is important to test the model's ability to generalize. This can be done by withholding a portion of the images from the training set and using them to test the model's ability to correctly identify the painter. If the model is able to accurately identify the painters based on images it has not seen before, it suggests that the distinct classes are representative of stylistic similarities. The first key challenge stemming from this, is the model's frame of reference. Everything it learns is framed through the class selection and has to be interpreted in relation to that. Therefore, selection of classes plays a major role in our approach, as elaborated in section 3.

23

Once verified through a model performance, the next step is to understand how the model is making its decisions. This process is often interpretative, as it involves inferring the decision-making processes from the classifications. However, there are also methods that provide insight into the model itself, for example with the use of heat maps [Selvaraju et al. 2017]; [Seifert et al. 2017]. Our focus lies not in explaining the decisions behind our model in detail, but the implications they have for our own analysis. For this purpose, the experiments have to be as transparent as possible from the beginning. We can achieve this by controlling our dataset and the receptive field of the classification task. Limiting and controlling the input, for example by omitting or including specific details, controls the model's decision-making from the beginning, rather than explaining it after training and validation. Since the painter attribution proves to be focused on the details of the paintings, laying a strong focus on them in the training process is not only preferable, but necessary.

24

This has been a key challenge for other projects as well. Some solutions include an automatic segmentation into tiles, as performed by Frank and Frank [Frank and Frank 2020]. They use the segments with the highest amount of information for their training. Another approach includes brushstroke-segmentation, which works particularly well for painters like van Gogh, who has a very distinct brushstroke-pattern [Li et al. 2011]. Both these approaches are not suitable for our goal of painter attribution. The lines of vase paintings do not contain repetitive patterns such as van Gogh. Additionally, we are looking for more complex combinations of brushstrokes which constitute the motifs: Comparing hands to hands, eyes to eyes. Other scholars, such as Bell and Offert [Bell and Offert 2021] and Langmead et al [Langmead et al. 2021] acknowledge this as well. This is why they use pre-existing models such as keypoint RCNN or the Google API for a semantic segmentation via object detection and face recognition. These pre-trained models are trained on real-world data however and do not work well with vase paintings. So instead of these approaches, we choose to annotate the images manually. The annotation strategy for this purpose is presented in section 4.

25

2.3 Functionality of Convolutional Neural Networks (CNNs)

Lastly, we want to give some technical background on the functionality of CNNs, because they also present us with certain characteristics to consider in our data compilation. CNNs have become the standard for image classification tasks in the recent years, with various models and model architectures being proposed (introductory works see [O'Shea and Nash 2015]; [Venkatesan and Li 2017]; [Khan et al. 2018]; [Li et al. 2021]). The success of AlexNet ([Krizhevsky et al. 2012]) and the introduction of ImageNet [Deng et al. 2009], the first large-scale annotated image dataset, have played a significant role in promoting the use of CNNs for

26

image classification.

Since then, various extensions have been proposed to enhance the representational power of convolutional architectures.^[7] At its core, a neural network is a complex mathematical model comprised of multiple layers. Each layer is a parameterized mathematical function. In the case of CNNs, the defining layer is the convolutional layer, where the mathematical operation of the same name is used. In this operation the input image is combined with a set of filters, resulting in feature maps. From a human point of view these feature maps look like a filtered image, with specific characteristics emphasized. By using different filters, the neural network can recognize patterns and determine them as features for the respective class. This is the same concept used in Gaussian blurring and edge detection. Using a blurring filter the CNN would recognize features regarding the overall shapes, while an edge detecting filter would emphasize the lines. By stacking multiple convolutional layers, the network can learn representations of increasing complexity. Convolutional layers are usually coupled with a pooling layer to condense the features in the feature maps, and an activation function to introduce non-linearity to the model.^[8]

27

When we train a neural network, our goal is to find the best parameters in the layers by making them fit our training data as closely as possible. We have a multitude of parameters across the layers, making it impossible to compute an exact solution. This is where backpropagation comes into play. These iterative technique helps us fine-tune the network's parameters, by gradually reducing prediction errors.^[9] During training, the neural network learns and optimizes the parameters, such as weights and biases, that can then be applied to new, unseen data. To put it simply, it basically involves backtracking through the model and optimizing the parameters to reduce errors in the classification.

28

The theoretical basis of CNNs has a major repercussions for our research: Everything in an image is fair game, which, as widely known, creates a perceptual bias in the model [Offert and Bell 2021]. Because the convolution is applied on the whole image, global properties are more dominant [Islam et al. 2021], making overall textures easier to learn than specific details. This is something we have to keep in mind throughout the workflow.

29

3. Painter Selection and Biases

3.1 Biases in the Image Selection

As the first step of our data compilation strategy, we want to address the class selection. Selecting the classes for our use case means choosing which painters we want to use as examples in our study. Furthermore, it also involves the question which paintings to choose and which painter attributions by which scholars to consider when assigning a specific painting to a class. Overall, our painter selection is influenced by both humanities and machine learning factors, with archaeological interests being the leading force. For the algorithms to work, some statistical requirements have to be considered as well:

30

First, the algorithms call for a sufficient number of samples per painter, which should also be in a similar range for all painters to avoid a skewed distribution. Second, we have to keep our dataset diverse enough in every aspect that may possibly be relevant, otherwise the model will not be able to generalize to unseen images. This means that we have to choose paintings representative for the variety within a class. Creating such a dataset is not a problem for most machine learning applications as long as they work with real-world data. It is always easy to produce more data to balance or enlarge the dataset. With vase paintings however, we 'only' have the preserved images and simply cannot create more. Projects in the Digital Humanities often deal with the bottleneck of too few or too diverse data. In other projects on vase paintings there have been attempts to enhance the datasets artificially, for example by transferring the style of vase paintings onto real-world images and therefore creating more images in the same style [Madhu et al. 2022]. Adrian Ryan even drew new paintings by his own hand [Ryan 2009]. However, we have refrained from artificially enhancing our dataset, because it usually introduces new diversity into the data which might create another bias. Also, our research question is focused on individual renderings, which cannot be re-created faithfully. So at least for the initial set-up, we had to focus on painters with a naturally large body of work.

31

However, this requirement raises another important question: By which criteria does a painting belong to a class? As discussed above, painter attribution on Attic vases is often done ambiguously and intuitively, making it difficult to objectify the attributions and put them into distinct classes. Beazley's attributions are so vast in number that they offer a solid basis for machine learning algorithms. Nonetheless, not all of his attributions are undisputed.^[10] The only factual evidence nobody can dispute are signatures. They are the real ground truth, even if we add the attributions of scholars who have a life-time of visual experience. Also, we do not want to limit ourselves to merely replicating Beazley's approach. We wanted to study connoisseurship and vase paintings in general. So, we also had to consider the number of signatures and potential controversies in regarding the painter's bodies of works to avoid too much disagreement within the classes. Nevertheless, we did include some controversial paintings in the dataset, because we wanted to study how they would perform. They could give us an insight into which features might make it hard for human scholars to agree on an attribution and therefore add to our understanding of the method.

32

Adding to these criteria, Attic vase paintings also confront us with inherent biases. They are a complex medium with dense information, pertaining to both formal characteristics and content. We are mostly interested in the formal characteristics, as interpretative information is not relevant for painter attribution. The categories can correlate however. To identify biases in this labyrinth of features, we defined different levels of formal characteristics, going from general and external to specific and deep. General characteristics are, for example, the composition of a painting, or its motif. These are dominant characteristics that take up a lot of literal space in the painting. A deep characteristic, for example, is the personal style, which is found within the details in specific areas of the painting. Such details might be overshadowed by more general characteristics and potentially be "buried" underneath them.

33

Above these levels of image-inherent characteristics, there is an even more general level. It concerns external criteria, such as the image properties and image quality: If the model is expected to perform on images with different angles, lighting conditions and other variations, then these factors need to be accounted for in the training data as well. So, the images we chose had to be diverse in this respect as well, which is further limited by the natural occurrence of vase painting publications, meaning printed books and online databases. We included scanned and de-rasterized photographs from printed publications, photographs on Wikimedia commons (<https://commons.wikimedia.org>) and photographs from major museum databases online, especially the British Museum in London (<https://www.britishmuseum.org/collection>), the Metropolitan Museum in New York (<https://www.metmuseum.org/art/the-collection>), the J. Paul Getty Museum in Los Angeles (<https://www.getty.edu/art/collection/>). While keeping the dataset diverse, it is also important not to oversample images with certain characteristics for a specific painter, as they contain easy-to-learn textural qualities. If we chose only medium quality scans of book publications for one painter and high-resolution camera pictures from Wikimedia for another, the model would generalize unwanted features. Thus, it is important to consider the overall quality within the classes carefully to ensure that it is representative of the dataset and free from external biases.

34

Going deeper into the levels of image information we can proceed from the more general to the more detailed levels. For a human it is easier to look for specific details, because we can control our visual attention. A neural network on the other hand, might get stuck on the more general shapes if they are generalizable,

35

due to the nature of convolution. This makes it even more urgent that we carefully consider these levels and their connections between each other. Sometimes, they are hard to separate, with the following characteristics being worth exploring further:

- The first characteristic concerns the context of the painting, meaning its placement and size. Vases offer many differently shaped surfaces for painting, which can affect the basic structure of the picture. Again, as with handwriting, the linework might vary depending on the available space and on the vessel's curvature. If a painter preferred certain shapes, this might lead to an unwanted bias.
- The next characteristics include motivic features, for example composition, scene, motif or the topic depicted. Those characteristics are often the first thing we notice in a picture, even if we reduce the depiction significantly. A neural network might learn those features before it learns anything else. If a painter always prefers certain motifs, the neural network might generalize this characteristic over the personal style, due to the space they take up in a painting.
- The final characteristic is found at the same "depth" as personal style: The *Zeitstil* (period or temporal style). As in all time periods, certain visual features are typical – "in fashion" – for their time. A good example of this is the clothing and the way the folds are painted. They are very distinctive in different time periods [Langlotz1920]. Both of these stylistic features are contained in the same details, which makes it hard to separate them.

Temporal style and personal style get confused sometimes and a study including different time periods might lead to a classification based on art periods or technical differences, rather than personal styles. The model might perform its task correctly at first glance, but the features it learns are not the ones intended. Still, the temporal bias is the easiest to resolve, at least in our initial set-up: By focusing on only one art period with its immediate predecessors and successors we can eradicate a potential temporal bias and teach the neural network to recognize subtle differences between personal styles instead. This approach was also considered in some of the projects mentioned above [Frank and Frank 2022, 840]; [Bell and Offert 2021].

The time frame we choose is the early 5th century BC, the late Archaic / early Classical times. The production of vase paintings in Athens spans over several centuries, with many stylistic and technical developments. Our chosen time frame takes place after the transition from black-figure to red-figure technique. In black-figure technique the figures were applied in a translucent slip onto the terracotta surface, which became black after firing the vessel. The details were then etched on with a sharp tool. Red-figure vases are the opposite: The vase was covered in black slip and the figures are blank areas that appear in the red colour of the terracotta. The contours and details are painted with a brush before firing the vase [Noble 1988, 99–117]. In our chosen time-period, the red-figure style had just become firmly established, while the generation before that is considered more experimental [Boardman 1975, 89–91]. It is also a time frame where it was common to sign the vases, which offers a good foundation for our dataset as well [Bolmarcich and Muskett 2017]; [Sapirstein 2013].

The vase size and shape, as well as the depicted motif, pose a challenge that cannot be solved this easily, because they may differ from painter to painter, even within a certain period. Between these two, size and shape are easier to control: They represent parameters that depend on the kind of vase onto which the image was painted on. According to Beazley, vase painters of that period can be roughly divided into cup painters and pot painters [Beazley 1961]. *Cups* in this terminology means smaller vessels from which one can drink, for example kylikes (Figure 3). They have a round, shallow body and are decorated on the curved outside as well as the plane inside. Meaning there are two kinds surfaces with different curvatures. But there are also other vases, such as skyphoi, that are closer in shape to the kind of cups we use today. *Pots* include large vessels, often with a large body, a small foot, and a shoulder connecting to the neck of the vessel. At the top they are either curved inwards, or curved outwards. The body is usually broad, offering a slightly curved surface for the paintings. However, the curvature at the top of the painting, and the size of the vessels differ greatly from pot-shape to pot-shape, as seen in the examples in Figure 4. There are also many exceptions, for example lekythoi, which have a narrow and high body, or decorative zones featuring smaller figures on the necks of larger vessels. Moreover, this distinction between cup and pot painters does not mean that they only painted these specific vessels. There are exceptions to the case, which adds to the variety within the classes. The cup painter Douris, for example, also painted a larger vessel, a psykter, as indicated by his signature (BAPD 205309, London, British Museum E768). Other pots can also be attributed to his body of work on a stylistic basis, such as an amphora in St. Petersburg (BAPD 205310, State Hermitage Museum B5576.) and another one in Paris (BAPD 205311, Louvre S3853).

Possible solutions against this bias would include focusing only on pot (or cup) painters or creating sub-classes per painter. Only focusing on one kind of painter would take too much complexity from the task, however, and due to the high variety and the uneven distribution of shapes, creating subclasses is impossible. This diversity of vase shapes causes a bias with unpredictable extent. Therefore, contrary to our approach with the temporal bias, we decided to include both pot and cup painters. We specifically included vessels representing exceptions to their usual preference of vase shapes, to emphasize the variation of linework in our dataset.



Figure 3. Cups by Makron. Top: BAPD 204734, New York Metropolitan Museum (MET) 06.1152; Bottom: BAPD 6917, MET 1979.11.9

A similar reasoning was used on the variety of the motifs. The depictions range from mythological fight scenes, to athletes doing sports, to women in their chamber. There is great variety in the topics depicted, but certain types of scenes or figures, such as mantle figures [Franceschini 2018], are often reused. As with the vase shapes, it seems that certain painters preferred certain motifs. The Berlin Painter for example frequently depicted musicians with stringed instruments ([Padgett (ed.) 2017, 4, 13, 15–16, 18–20, 28, 33, 39, 55]. This observation can lead to circular reasoning, with attributions being based on the motifs. Since the dataset is limited, we cannot simply avoid having too many examples of a painter's favourite motifs. We can only ensure as much diversity as possible per class. Another strategy against this bias consists of stripping the details from their compositional and motivic context by segmentation, as explained in Section 4.

40

3.2 Selection of Painters and their Circles

By reducing these biases, our data selection was shaped in a certain way, either by focusing on certain characteristic or by ensuring diversity so that the model would be able to generalize. The final selection was further determined by painters' potential for answering research questions and uncovering similarities and connections between them. With all this in mind, we chose four painters from around 500 to 470 BC: The **Berlin Painter** [Beazley 1963]; [Padgett (ed.) 2017], **Douris** [Beazley 1963, 425ff]; [Buitron-Oliver 1995], **Makron** [Beazley 1963, 458ff]; [Kunisch 1997] and the **Brygos Painter** [Beazley 1963]; [Pace 2019]. Additionally, we also added **Oltos** [Beazley 1963]; [Bruhn 1943]; [Hoffmeyer 1943]. He was active a little earlier than the other painters and painted many cups, but also other vase shapes. His performance in training and evaluation should give us a deeper insight into the two potential temporal biases and vase shape biases throughout the experiments.

41



Figure 4. Pots by the Berlin Painter. Top: Neckamphora: BAPD 201878, MET 07.286.69; BAPD 201837, Los Angeles Getty Museum of Art 86.AE.187; BAPD 201888, MET 41.162.17; Amphora: BAPD 201811, MET 56.171.38; Bottom: Stamnos: BAPD 201960, MET 1988.40; Hydria: BAPD 201987, MET 10.210.19; Oinochoe: BAPD 202004, MET 22.139.32; Lekythos: BAPD 202022, MET 21.88.163.

Three of these painters - the Berlin Painter, the Brygos Painter and Olto - were put into relation to other painters by different scholars. Sometimes, when talking how painters might relate to each other, scholars have been overly interpretative with their assertions based on visual similarity. To avoid misunderstandings regarding the interpretation of our data in relation to the history of research in Classical Archaeology, we have to be very careful in our choice of words. Firstly, we will use the word 'relationship' only when talking about clearly definable relationships between painters. To this, we add the terms "relate to" and "connection". With connection we mean visual similarity, which includes the possibility that there might not be a historical relationship at all. Second, we call the three painters "core painters" (= training classes) and will describe their groups of proximal painters as a "circles" (= test classes). Beazley used the word "circle" to describe an extended range of relationships beyond workshops or groups [Robertson 1989]. In contrast, we use this term to include all kinds of relation(ship)s and connections that archaeological research claimed between certain painters and their styles of drawing. We neither agree nor disagree with this research and its interpretations. We simply use these circles to explore and further define the visual relations they are based on.

42

The three circles are based on different dimensions of argumentation. The first is the circle around Olto. He signed his works frequently and many of his vases were also signed by their potters, which was used before to form a network of workshop relationships [Cline and Hasaki 2019]. The second is the circle of the Brygos Painter. We do not know his real name, because there are no signed works. However, Beazley believed that he potentially worked together with a group of stylistically similar painters [Beazley 1963, 368, 400–405]. His assertions are supported by a potter debris found in Athens with fragments of some of the visually similar painters [Tzachou-Alexandri 1968]; [Baziotopoulou-Valavani 1994]; [Monaco 2000]; [Maffre 2001]; [Williams 2017, 168]. The fragments are also unsigned and it is therefore important to be wary of circular reasoning when discussing this group. Both Olto and Brygos may give us insight into workshop processes and how the paintings of fellow craftsmen might relate to each other stylistically.

43

The third circle surrounds the Berlin Painter and is based on purely stylistic observations. His circle is a prime example of historical interpretations of visual similarity, because it contains what connoisseurs label "teacher and student relationships." It is impossible to reconstruct such relationships based on straightforward evidence, instead it is based on the similarities between the painter's styles and vase shapes [Oakley 1997, 97]. Hypotheses such as these have been criticized in Classical Archaeology, because the arguments are circumstantial and too much historical information is inferred from visual observations [Neer 2009, 45]. Nevertheless, we want to study the Berlin Painter's circle in particular, because it represents the peak of connoisseurship intuition and association: Connoisseurs have seen distinct and nuanced similarities between the paintings that led them to label them with this specific interpretation. Our goal is to strip out all the historical assumptions and to re-examine said similarities. We will continue to use quotation marks when discussing their relationship, since they are only inferred.

44

	Supposed Relationship		Instances
Berlin P.	Core Painter		230
Providence P.	"Student"		84
Achilles P.	"Student"		164
Hermonax	"Student"		97
Phiale P.	"Student" of the Achilles P.		150
Oinokles	"Follower" of the Providence P.		48
Phintias	"Teacher"		43
Euthymides	"Teacher"		40
Kleophrades P.	"Student" of Phintias and Euthymides		305
Myson	"Student" of Phintias		42
Persephone P.	"akin in spirit" to the Achilles Painter		35
Additional Painters to diversify the data-set			
Douris			208
Makron			314
	Shared Potters		Instances
Oltos	Core Painter		150
Euphronios	2 (4 Signed, 8 Attributed)		94
Epiktetos	2 (4 Signed, 3 Attributed)		119
Onesimos	1 (1 Signed, 9 Attributed)		129
Nikosthenes	2 (15 Attributed)		31
Pistoxenos	1 (1 Attributed)		28
	Supposed Relationship	Fragments	Instances
Brygos P.	Core Painter	✓ (14)	197
Briseis P.	"Mild-Brygan-Group"	✓ (6)	100
Dokimasia P.	"Mild-Brygan-Group"	✓ (3)	62
Foundry P.	Visually Related by Beazley	? (1)	120
P. of the Paris G.	Visually Related by Beazley	? (1)	75
Onesimos	"Teacher" of the Foundry P.	? (1)	129

Table 1. Painters in our data-set and number of instances. The arguments for their inclusion are based on a literature of supposed relations. For the Brygos P., this includes fragments in the archaeological find and their attribution, for the Berlin Painter the labels given by research and for Oltos the number of signatures on the works by the same potters.

Table 1 shows the distribution of instances per painter in our dataset. An instance is a single figure in a painting. Asking “who painted this figure?” instead of “who painted this vase?” has two advantages: First, it allows us to disregard the overall image composition, therefore reducing the potential motif bias; and second, it leads to a more nuanced assessment of painter attribution and the role individual figures may play when the model attributes a painting.

45

In total, we carefully curated a dataset consisting of 2785 individual figures extracted from 1353 vase paintings created by the five core painters along with 19 other painters in their associated circles. Our total dataset includes 37 painters, 9,362 figures, and a total of 428,960 annotated objects. The distribution displays a natural imbalance, with a significant variation in the number of existing paintings across different painters. This inherent imbalance poses potential challenges for the training process, particularly when it means that classes are underrepresented because they have a low number of paintings. Such underrepresentation may adversely affect the model's performance, especially when assessing these painters during the testing phase. Fortunately, there are various methods to improve training when dealing with an imbalanced data source, which we will describe in Section 6 below.

46

4. Image Annotation and Segmentation

With the painters selected we can now move onto the input-segments. Segmenting the images means that our model is not trained on the single figure, but on the (combination of) details per figure. As discussed above, the established methods for automating semantic segmentation do not work for vase paintings. Instead, we opted for a semi-automated workflow with manually annotated images as the basis for our own object-detection model. Adding to the need of transparency, our strategy also aims at reducing the biases of motif and vase shape, thereby focusing on the details of a painting.

47

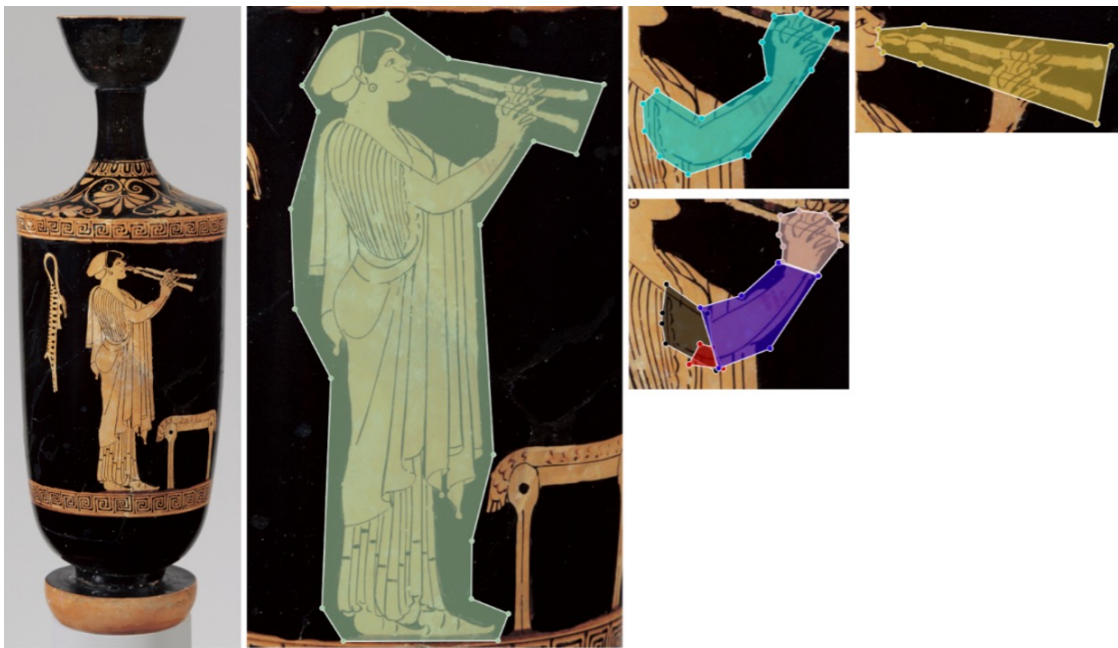


Figure 5. The annotation process involves annotating the figures at different levels of detail. The parts the figures are connected via parent-child-connections (for example the flute belongs to the hand, the hand to the arm, the arm to the woman). Vase depicted: BAPD 204117, MET 24.97.28

For this purpose, we have developed annotation guidelines with a controlled vocabulary. It was important to have an agreement between the annotators on what exactly the labels represent and which part of the painting they encompass. It is also important that they do justice to the painting's complexity and can be used not only our project, but also for future research. This requires detailed, comprehensive and explicit annotation guidelines, which we developed during our project. The guidelines include the following three core principles:

48

1. We want to compare different levels of detail, in which each label has to be connected to their corresponding higher-level-label (for example, an eye belongs to the head, a hand to an arm). For this, adapted a **hierarchical system**, in which the annotations get more detailed with each successive level, while the segments stay connected via parent-child connections. Specifically, this means that when we annotate a figure, we begin with the highest level: The figure itself. This can be either a man, a youth, a woman or a creature. Then we annotate the main body parts of the figure (for example the head or the arm). These body parts then get divided into their sub-parts. For the arm this would be the upper arm, the forearm, the elbow and the hand (s. Figure 3) The same goes for clothes: First, the specific garment is named; then the drape of the hem folds and the surface folds. This way we have three levels of detail: The figure, the larger body parts and the smaller details.
2. For this purpose, we need a controlled vocabulary with a focus on the formal characteristics of a painting. A neural network is unable to *understand* the image. It learns from our categories which therefore must be as clear as possible. This is an important consideration for our controlled vocabulary. On the one hand, it is composed of specific terms. These are usually the names of the body parts, but sometimes, if necessary, we used technical terms from Classical Archaeology, for example to describe the clothes. This is because the specific types of clothing come with the same folds within their class. A chiton has fine folds at the bottom, a himation has large folds. On the other hand, our vocabulary also goes beyond simply determining semantic segments: It also has a **visual component**, describing the composition of brushstrokes associated with each label. The hip, for example, consists of two lines going from the lower side of a figure to the genitals. The shoulder (Figure 6) is described by the concave shape from the neck to the upper arm. The goal of the vocabulary is to clearly define the combinations of brushstrokes associated with each segment.
3. Still, some segments vary depending on the context they of their depiction, even if they are clearly defined. Therefore, we had to include **additional information concerning the variation of the labels**. This additional information had to follow certain rules as well, to avoid overlapping and mix-ups between the strict form and levels of additional information. For example, we annotated whether or not a figure or their body parts are clothed, because the folds of the clothing change the depiction in a specific way. We also added information on the action the figure or their body parts are performing. A closed hand is a different shape than an open hand, for instance. We also added some interpretative information, in case that it correlates with specific shapes, such as the iconographical identification of gods or heroes. It is usually based on the figure's attributes, and thus contains information about its shape: Athena usually wears a large helmet, Eros has wings, Herakles wears a lion's skin, and so on. This way, a figure is identifiable using a collection of features within the image, which adds to the identification of a potential motif-bias.

Most of the annotated information was not specifically used for training, but for describing the shapes contained in an image beyond that. It also opens up the possibility for motif-dependent de-biasing both before and after training, for example if we decided to exclude all hands playing instruments from the dataset. This way we can exclude motif-heavy labels from the beginning, reducing the bias efficiently. We also wanted to pave the way for future research on the contents of the image, such as iconographical analysis, which greatly profits from such a detailed and universal annotation.

49

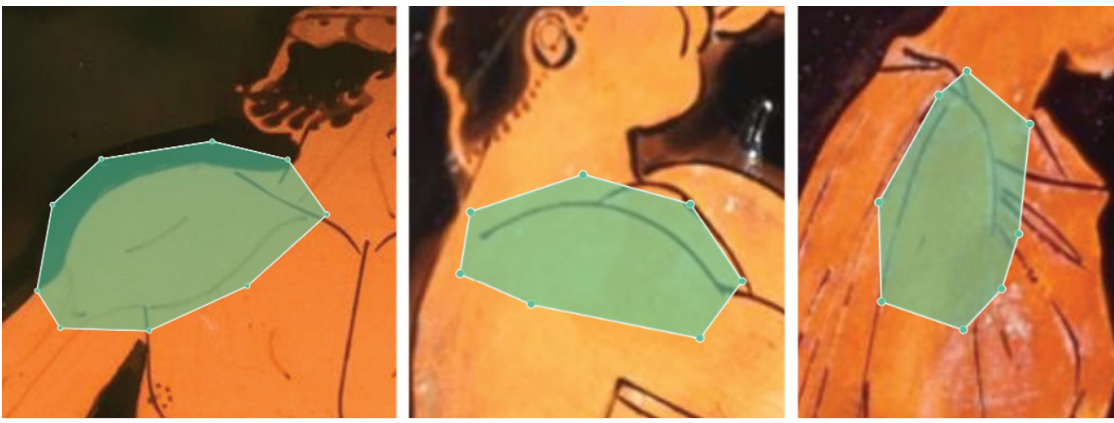


Figure 6. Annotation of the shoulder. Our polygons encompass clearly defined accumulations of brushstrokes, which do not always align with anatomical definitions of body parts, but rather the domain specific depiction thereof. From left to right: BAPD 201878, MET 07.286.69; BAPD 204497, MET 06.1021.188; BAPD 213926, MET 17.230.13.

The annotation team consisted of student assistants with a background in Classical Archaeology. To avoid an annotator's bias, we also split each painter between them. After annotating a large amount of images we used the manual annotations to implement our own object-detection model, which was trained to annotate the most common labels and also their the parent-child connections. We used our hierarchical system to make the process even more efficient: Our annotations are not nearly as numerous as real-world data. We resolved this problem by adding a control instance through a human annotator and initiating the detection in the same hierarchical order as our annotations: firstly, the human annotator selects the figure, which becomes the area for the regions-of-interest of the first hierarchical level (body parts). The process is then repeated for the sub-parts. This means, that the next level of annotation is only being searched for in the relevant regions, e. g. only looking for hands in the area of the arm. This made our detection very accurate despite the smaller training set [Kipke et al. 2022].

50

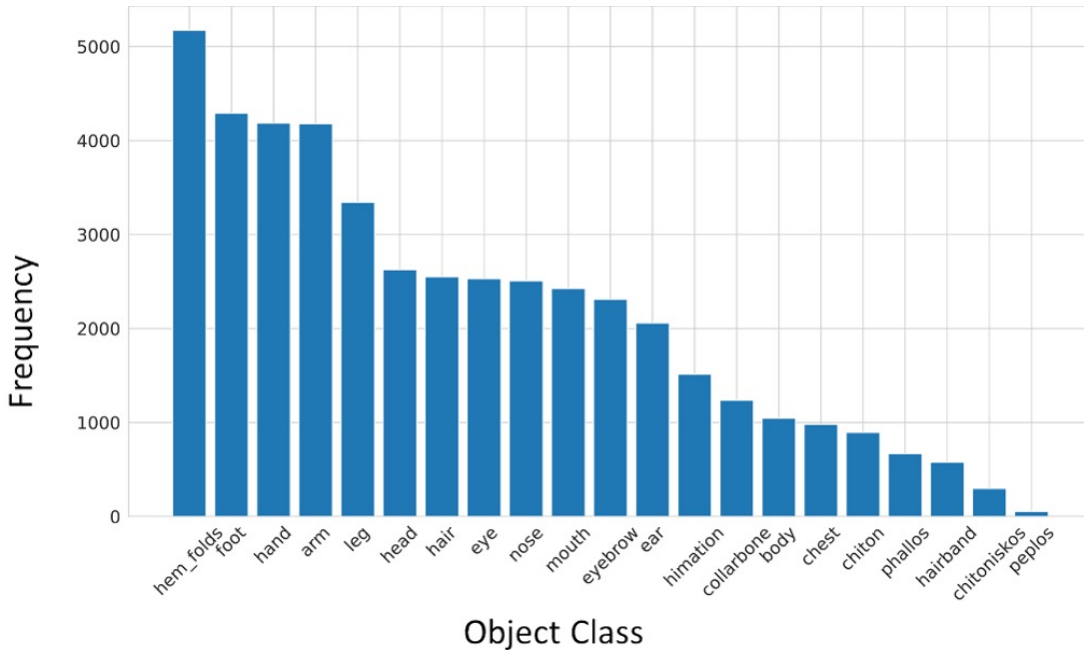


Figure 7. Frequency of the object classes for the segments extracted from each figure

For training and analysis, we used only the 21 most common object classes. We excluded the surroundings of the figures, as well as any objects they hold in their hands. Not only are those instances few, they also strongly indicate the motif of the image. In the end this annotation process resulted in a total of 45,442 components recorded across the paintings included in the experiments. This results in 16.3 ± 4.4 segments per figure (Figure 5).

51

5. Interim Conclusion: What goes into the neural network?

In the end it comes down to the question of what we want the network to learn and which biases we plant into the training data, whether unknowingly or deliberately. This challenge begins with the choice of painters, continues in the selection of images and culminates in the segmentation of the individual images. In our selection we aimed at balancing potential biases using different strategies:

52

1. **Limiting** the painter selection to a specific characteristic (the training classes represent only two generations, which helped us to avoid a temporal bias)
2. **Maintaining** heterogeneity with respect to the characteristics affected by potential biases on the same level as individual style, so as not to reduce the complexity of the task.
3. **Reducing biases** concerning full image characteristics (such as motif or composition) by **segmenting the image and excluding labels** that carry

characteristics of this bias (attributes, surroundings).

It boils down to a question of representability. We want our data-set to represent the variation of the linework, while ignoring other image characteristics. Our decisions are based on archaeological criteria, but with machine learning algorithms in mind. This means that our data already underwent several layers of processing, which differentiates them from the original pictures a Classical archaeologist might look at:

53

1. The very photography of the vase is the first layer of distortion. By publishing vases as photographs, we are transforming the paintings from a three-dimensional surface into two dimensions, which results in details being distorted depending on the angle from which the picture was taken.
2. By sorting and compiling the data we contextualize the images in such a way, that the neural network has to assume common features within a class. This creates a framework, in which it can only "see" generalizable features of the images.
3. Our image annotation cuts each image into pieces. This way, we try to control which aspects add to the training of the model. We are trying to re-create the way an archaeologist is able to focus on certain details, even if it is just an approximation and not possible for every detail to the same extent.

6. Experimental Set-Up

6.1 Experimental Details

With the data sampling laid down, we now come to our machine learning experiments. First, we want to elaborate on some of the experimental details, because they set the baseline for the analysis down the line. Addressing the imbalanced nature of our dataset, where the number of figures per painter varies drastically, we also employed standard data augmentation and overfitting-avoidance strategies. ^[11]

54

To assess the performance of our models, we employed a 10-fold cross-validation methodology [Berrar et al. 2019]. This is a method to make use of the data in smaller datasets as efficiently as possible. In this approach, we divided the images of each painter into ten equal parts (folds). For each of the ten experiments, one fold from each painter was assigned to the test set, while the remaining data was used for training and validation. This way different parts of the dataset rotated through the training / validation / test split and added to the model's ability to generalize. This means, that looking at each fold we get an insight in how well an instance performs not only as training, but also as test and validation data.

55

We also employed transfer learning techniques. In many use cases, the neural network is not trained from scratch. In the lower layers it learns basic shapes and edges, which stay the same between different data domains. That is why many projects rely on transfer learning, where a pre-trained neural network is adapted to another data domain. In our work we employed the large feature extractor ResNet-50, which is pre-trained on ImageNet [He, Zhang, Ren, and Sun 2016]. ResNet-50 is the appropriate size for our experiments, because smaller versions resulted in a loss of performance and larger versions did not offer a significant lift. On top of the ResNet-50 model, we added layers with our own classification task. The pre-trained model is subject to its own data biases, through the adaption to our own data domain they do become mostly irrelevant however.

56

6.2 Image and Image Set Classification

As discussed above, there are many arguments in favour of a model trained on the details of the paintings. However, to confirm the efficacy of this approach, we trained models on other potential scenarios as well. In this experiment, we distinguish three scenarios, each of which gave us unique insights into the process. These approaches differ in their methodology and underlying assumptions, providing us with a comprehensive understanding of the task at hand. Figure 9 shows examples of the three scenarios:

57

1. The first approach involves a prediction based on the complete vase painting images. This encompasses not only the figures but also the overall scene, motif, ornaments and parts of the vase. We wanted to include the biases discussed above to analyse how they affect the model.
2. In the second approach, our focus shifted to detecting the painter based on a single figure extracted from the entire painting. This approach aims to eliminate some of the biases such as figure's surroundings, ornaments etc. We wanted to see what remains and how much this approach differs from (1) and (3), since it borrowed from both, but has the advantages of neither.
3. The third approach constitutes our preferred strategy. Here, instead of presenting the model with the figure as a single image, we represented it as a collection of its sub-components, such as the arm, hand, and eye.

This meant that we worked not only with different data inputs, but also with different classification set-ups. While the first two scenarios follow standard image classification protocols ^[12], the third scenario comes with the additional challenge of controlling the segments. The complexity of this task calls for an extension of the standard set-up to handle a varying number of images, with each image representing a segment of the original painting. While the overall objective remained the same, our model now had to process sets of images, while maintaining invariance to the order and number of elements in each set.

58

Recent studies have investigated the concept of invariant and equivariant layers from various perspectives. Maron et al. [Maron et al. 2018] introduced a methodology that involves a set of learned basis functions, while the set-transformer approach [Lee et al. 2019] employs straightforward self-attention mechanisms. Zaheer et al. proposed equivariant and invariant layers capable of operating on a set of elements by building upon the Kolmogorov-Arnold representation theorem as discussed by [Zaheer 2017]. This theorem asserts that any multivariate continuous function ϕ can be expressed as a finite composition of continuous functions of a single variable and the binary operation of addition. ^[13] This means specifically, that our network consists of two components. The first component functions as an inner function f . After aggregating the images through summation, a second model g consisting of four dense layers is employed to compute the painter prediction for the entire set. Thus, the networks f and g are trained on the precomputed embeddings of our paintings to classify the actual painter. In this study, we have chosen to utilize a deep sets architecture due to its inherent simplicity, which makes it suitable for addressing problems characterized by limited available data. ^[14]

59

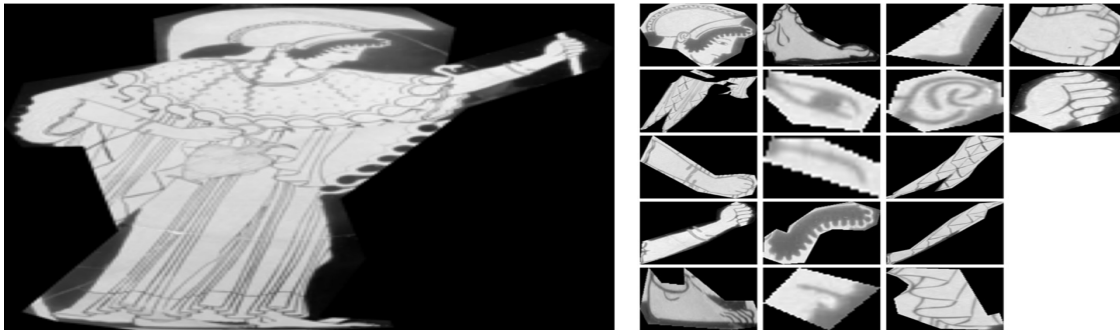


Figure 8. Example of the three different scenarios for processing the full painting, a single figure, and the set of annotations. Vase depicted: BAPD 201960, MET 1988.40

60

It is also important to note that the experiments involving the full figure and the figure set employed the same split of figures, while the full image scenario involved a different random split across all paintings. Additionally, we ensured that figures or paintings from the same vase were not present in both the training and test sets simultaneously. Apart from these distinctions, the general experimental set-up and pre-processing across these three variants remained consistent.

6.3 Pre-Processing and Image Distortion

61

Another point to address is the input distortion. Not only are the images scaled down for training, they have to be normalized to a fixed aspect ratio as well. Since we relied on a pre-trained deep feature extractor, the size of the input image is already predetermined to be 224 x 224 px by the pretraining dataset, even if we annotated high resolution images. Operating on full high-resolution images, which can be up to 4000 x 4000 px, would limit the number of images we can process simultaneously, potentially slowing or even impairing the training process.

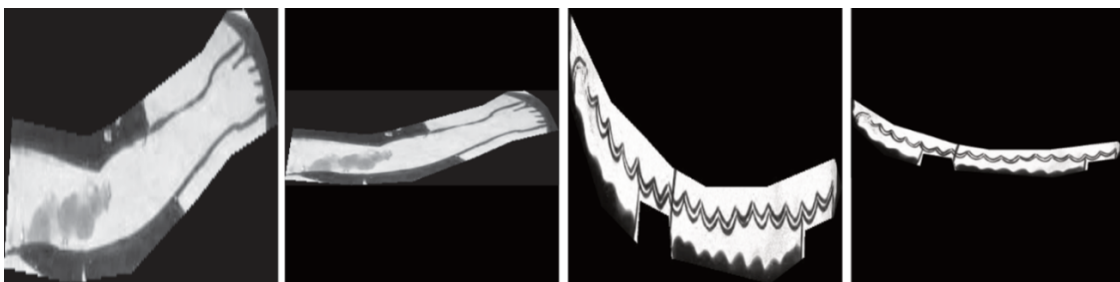


Figure 9. The arm and hem folds scaled (left) and padded (right). With scaling the linework gets distorted significantly, especially for narrow and long labels. Left: BAPD 200751, MET 09.221.47; Right: BAPD 201811, MET 56.171.38

62

Although these methods are advantageous due to their simplicity and low computational cost, they have the drawback that they distort the image significantly. Every segment is scaled without respect to their ratio, making small details larger and vertically compressing labels such as the arms or hem folds (Figure 8). While this distortion is considered negligible for standard computer vision tasks [Hashemmi 2019], it was crucial to investigate this factor in our research, since the specific linework observed is assumed to have a significant impact on painter attribution. Does the contortion pose a disadvantage for the task or is the model able to compensate for it?

Accuracy with standard deviation training with....		
	...paded images	...scaled images
Berlin P.	61.7 ± 5.1	71.3 ± 4.2
Brygos P.	54.3 ± 5.9	53.3 ± 5.6
Douris	50.8 ± 8.7	62.1 ± 7.5
Makron	66.5 ± 8.9	78.9 ± 5.2
Oltos	56.1 ± 2.8	54.7 ± 9.1
Average	58.9 ± 2.9	66.3 ± 3.4

Table 2. Comparison of the performance between padded images and scaled images. Results are given as accuracy aggregated across the 10-fold cross-validation

We tested this by training three models on different pre-processing scenarios. In the first variation we kept the distorting normalisation, in the second scenario we scaled the image to their longest side and added black padding to the empty space around it (Figure 8). In the third variation we refrained from up-scaling smaller segments, as this results in different line thicknesses for different labels depending on their size. As expected, however, the third scenario did not create satisfactory results due to the small size. The comparison between the first two other scenarios can be found in table 2. Contrary to our concerns, the distorted images do in fact result in higher scores. This may be because scaling the images maximizes the input information, while padding the images adds large areas of black background, which results in unused space. In the case of the hem-folds and the arm the distortion might even be helpful. For the hem-folds it emphasizes the linework and rhythm of the folds, while for the arms, the focus is on the changes in direction of the lines rather than their length.

63

Additionally to the input-distortion, the images are often naturally distorted due to their curvature. Especially the hydria in Figure 4 shows clearly, how some paintings are difficult to photograph. This difficulty is specific to certain vase shapes and brings us back to the problem of a potential vessel shape bias. The problem of vase painting photography has been addressed on other occasions and with un-warping-algorithms (f. e. [Felicisimo et al. 2019]). However, they add another level of distortion in the pre-processing level, skewing the dataset even further. Due to the heterogeneity of our image selection concerning vase shapes and the focus on details rather than the whole vase, it would be impossible to create a workflow that addressed all of them. The findings of the padding vs. scaling experiment support our decision, as distortion in itself does not seem to affect the outcome of the experiment negatively, except for a potential shape bias. However, this should of course continue to be tested in further research and the methods should be adjusted if necessary.

64

6.4 Robustness

Finally, we also want to address model robustness. "Robustness", in this context, refers to the stability and consistency of a model's predictions when trained multiple times. Neural networks exhibit both determinism and non-determinism in their behaviour. When fully trained, (non-bayesian) neural networks operate deterministically, consistently producing the same outputs for a given input, ensuring their reliability. However, since the parameters of a neural network are typically initialized randomly before training, model performance might vary across multiple training runs, even with the same architecture and data. It is important to quantify this variation to ensure that the results hold ground, not only regarding the robustness of classes, but also the single instances.

65

	Accuracy	Standard deviation over...	
		...folds	...runs
Berlin P.	73.9	± 4.2	± 1.5
Brygos P.	61.5	± 5.6	± 3.0
Douris	58.6	± 7.5	± 3.2
Makron	79.3	± 5.2	± 1.9
Oltos	65.3	± 9.1	± 2.5
Average	67.7	± 6.32	± 2.42

Table 3. Standard deviation over the test-training folds and over then re-trainings with the same set-up

	Most frequent class	Second frequent class	Third frequent class
Berlin P.	Berlin P. (73.9%)	Brygos P. (11.4%)	Makron (8.1%)
Brygos P.	Brygos P. (61.5%)	Makron (9.9%)	Douris (8.7%)
Douris	Douris (58.6%)	Makron (13.8)	Brygos (11.3%)
Makron	Makron (79.3%)	Brygos P. (14.7%)	Berlin (7.1%)
Oltos	Oltos (65.3%)	Berlin (7.8%)	Brygos (7.0%)

Table 4. Classification accuracy of each painter along with the 2nd and 3rd most frequently assigned classes

This investigation primarily focused on two key dimensions: First, the ability of the model's predictions to withstand repeated executions on identical but permuted inputs; and second, the possible alignment of a high confidence level in the model's class scores with the robustness of its predictions. For this purpose, we repeated our main experiment with the same 10-fold cross-validation image sets 10 times. We then analysed the specific class assignments across these 10 runs. As shown in table 3, the accuracy over runs varies only up to ± 2.42%, meaning our best model achieves an overall accuracy of 70.12% ± standard deviation.

66

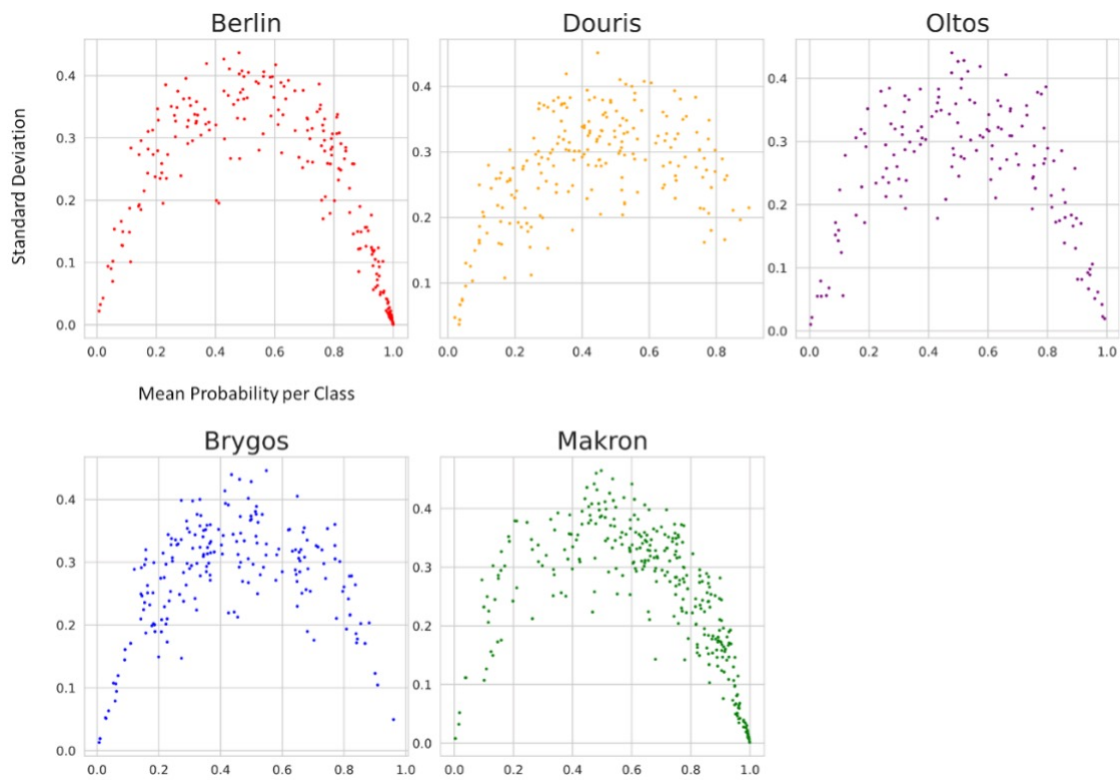


Figure 10. Class distribution over 10 repeated training runs with random weight initialization. The scatter plots show the relation between mean assigned class probability for the correct class of an instance and its standard deviation

Two more experiments give us a better insight into the different aspects of the robustness. First, we examined the instances for each painter and analysed the correlation of the standard deviation and prediction value. The prediction value is the raw model output and represents the level of certainty for the attribution of an instance. The scatter plot in Figure 10 shows the mean of the assigned class probability over the 10 repeated runs for each figure, plotted against the standard deviation over these runs. It shows a curve, in which a high mean probability corresponds with a low standard deviation. Similarly, a low score also corresponds with a low standard deviation, meaning strong predictions come with a high confidence and predictions in the middle field tend to vary over the runs. This suggests that the predictions with a very high, but also with very low prediction values are especially stable, meaning that they might provide good case studies for studying painter attribution.

67

This became even more apparent looking at each assigned class ordered by its frequency. For this, we analysed the class prediction with the highest probability and evaluated the frequencies over assigned class: Did the model assigned the same class 10 out of 10 times, or did the final prediction change frequently? The boxplot in Figure 11 shows these frequencies. The mean/centre of each boxplot corresponds to the relative frequency over predicted classes. Over 10 runs, the model would assign the most frequent class for an instance 70% of the time; conversely, the second most frequent class was assigned at a rate of only 20%.

68

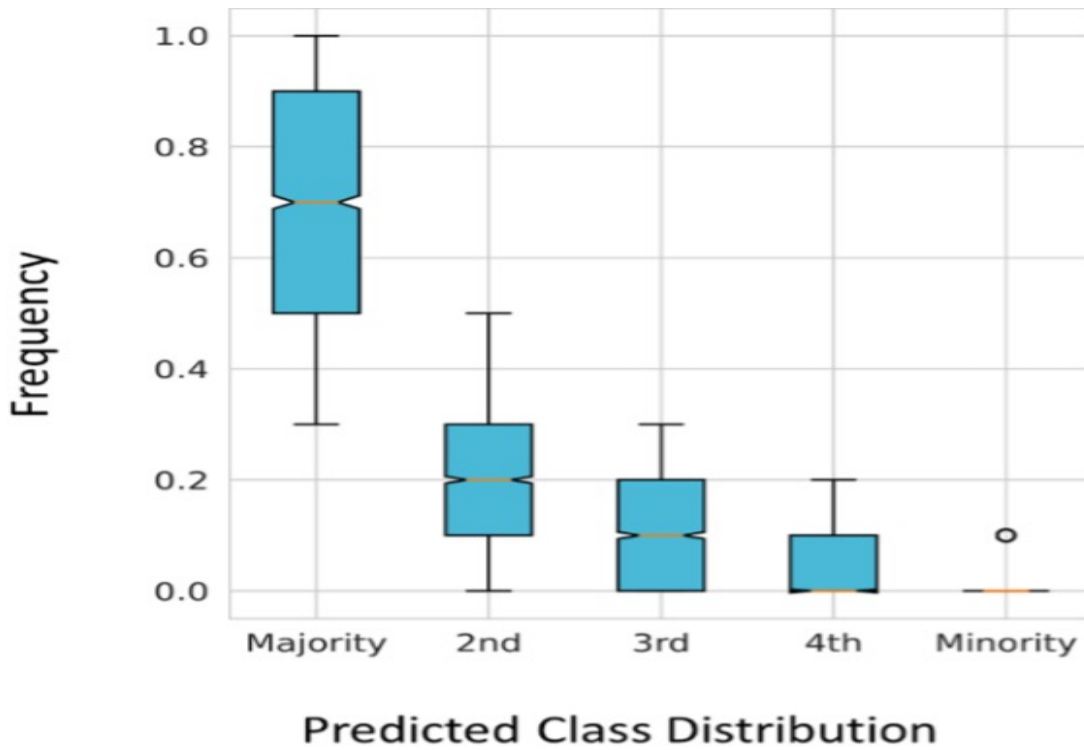


Figure 11. Class distribution over 10 repeated training runs with random weight initialization. Left to right shows the boxplot over the relative frequency of the class assigned the most, then second most, etc., up the class least likely to be assigned.

This means that the model is fairly sure about the overall classification, even when it misclassifies an instance and, in that regard, it is stable when making attributions, even if they are wrong. In further analysis, the second highest class might still be relevant, as it attests which painters get confused more easily than others. The second and third frequent classes can be found in table 4. Overall, these experiments add an important perspective for the analysis of peculiar attributions, especially if they are confidently wrong.

69

7. A First Look at the Results

7.1 Image and Image-Set Classification in Comparison

Finally, we turn to the results of our experiments, starting with the comparison of training scenarios regarding the input images (table 5). The difference in accuracy between full vase paintings segment sets is surprisingly small. Considering the fact that the data is highly complex and that the task requires deep and specific knowledge (in human experts as well) the numbers up to 70% accuracy with 5 classes is within expected parameters. The full figures have an accuracy about 17% lower than the other scenario.

70

	(1) Full Image	(2) Single Figure	(3) Segment-Sets
Berlin P.	86.6 ± 6.6	56.9 ± 2.7	73.9 ± 4.2
Brygos P.	56.6 ± 5.5	47.2 ± 3.6	61.5 ± 5.6
Douris	55.7 ± 3.0	42.8 ± 5.1	58.6 ± 7.5
Makron	73.7 ± 3.7	63.4 ± 2.9	79.3 ± 5.2
Oltos	66.2 ± 5.8	42.1 ± 8.6	65.3 ± 9.1
Average	67.9 ± 4.9	50.5 ± 4.5	67.7 ± 6.32

Table 5. Accuracy with standard deviation of each scenario described above in Section 5.2. Results are given as accuracy aggregated across 10 folds.

When comparing the numbers, we have to be mindful of the major differences in scaling. In the case of scenarios (1) and (2), most elements within the figure, like the eyes, were exceedingly small. This, in turn, means that the linework of the details cannot be a factor for attribution. The difference between full vase paintings and the other two scenarios demonstrates the impact of various factors beyond the distinctive linework. This includes elements such as the overall composition of the painting, the placement of figures and the vessel shape. Because this confirms that the model is able to generalize other characteristics of a painting between the painters and make correct attributions on that basis. In comparison to the traditional method, this is not inherently misleading. Archaeological scholars do not disregard composition and arrangement when making painter attributions either. Beazley for example emphasizes aspects of composition [Beazley 1930, 12] and the overall impression of the figures [Beazley 1933, 7]. This does, however, mean, that we are restrained to only general characteristics with this model.

71

It is also noticeable, that the accuracies vary significantly between the classes. Some are easier or harder to learn in either of the scenarios, others only pertaining to their full images or single figures. Douris and Makron have a higher accuracy with segment sets, which is not surprising, considering their datasets included mostly the outsides of the cups as well as the insides. In terms of vase shape and composition the classes have as much in common with each other as they do

72

within the class. Therefore, they are more distinguishable by their details. The Berlin Painter, on the other hand, has a higher accuracy for full images, which indicates, that his class might be more coherent in composition and motif.

The second scenario, which involves only full figures, combines the weaknesses of the other scenarios: Cutting out the figures reduced the stronger foreground characteristics of (1) and focusing on the whole figure makes it impossible to recognize details such as in (3). Therefore, the significant drop in accuracy is not surprising: Neither the overall characteristics of the image nor the details of the handwriting help the model to recognize the painters. This is an important observation: The biases apparent in (1) are significantly reduced when focusing on a figure. Thus, we feel validated in focusing on the details of the figures, which leads to a significant performance enhancement in (3). This can be attributed to the model's ability to process the information in the details at a much higher resolution, leading to a superior recognition of individual lines and details neglected in the other two scenarios.

In addition to the accuracy, we also took the standard deviation into account, which is slightly larger in (3) than in (1), especially for specific painters. The standard deviation is calculated across the cross-validation, meaning that a higher standard deviation indicates that the model struggled to generalize across the varying training and validation data. Oltos was exceptional in this sense, with a standard deviation of ± 9.1 . There are three reasons for this: First, he is the painter with the least instances, making him prone to misclassification. The difference is not very high however and there are more reasons found in the composition of his oeuvre. Therefore: Second, his oeuvre is quite heterogenous, making it difficult to generalize learned features across the images. And third, he is the earliest painter. The models difficulties with him confirm our hypothesis regarding the model's focus on the personal style of painters from our main period. Oltos stands on a different level, as he should. Another painter with a higher standard deviation is Douris (7.5). His case is more difficult to explain. His work is often divided into distinct phases (Buitron-Oliver 1995 2-3). Arguments vary from figural style to his specific use of ornaments. It is possible that his class is more heterogeneous and that his development therefore posed a challenge for our model.

This experiment supports us in our proposed approach in two aspects: First, it corroborated the assessment of potential biases and the strategies to reduce them. When we compared the performance between (1) and (3) we noticed that our approach works well, specifically for the cup painters. Second, the combination of details allows the model to capture both larger elements and smaller details in a much more nuanced manner. It also proves to be invaluable for our aim of understanding the specific criteria contributing to each painter's attribution, and aligns with our primary goal of challenging and questioning established methods. However, it also indicates that the full image classification might provide an additional insight into our dataset composition and possible biases in further experiments. Therefore, it is a valuable addition, even though we are aware that this is not a recognition of 'style' in the sense of handwriting.

7.2 Attribution of Painter Circles

Next, we want to take a first look into the attribution of the core painter's circles. Since our model has only been trained to differentiate between the five core painters, testing it on other painters gives us an insight into their stylistic proximity to the training classes. For an additional perspective on the level of similarity we considered the full image model as well as the model trained on segment sets (table 6).

First, the results vary significantly for the respective circles. As explained in Section 3, there are different underlying assumptions for each circle. The model worked best for the Berlin Painter, whose circle is defined by stylistic similarities and supposed "teacher-student-relationships". In the Brygos Painter's circle, the model classified some painters relatively well, especially the Briseis painter [Beazley 1963, 406] and the Dokimasia painter [Beazley 1963, 412] who are considered to be a "milder" variation of the Brygos Painter [Beazley 1963, 400]; [Giudice and Giudice 2011]; [Williams 2017, 166]. The Foundry painter and the painter of the Paris Gigantomachy both have high prediction values in full figure, but not in segment-set-classification. Beazley described the style of the Foundry painter as follows: "*with his forcible, sometimes even brutal style he often equals the Brygos Painter*" and the Painter of the Paris Gigantomachy "*stands nearer to the Foundry Painter than to the Brygos Painter himself*" [Beazley 1963, 400]. It may be that our model classifies them as less similar to the Brygos Painter because the connection is based on the general impression more so than on their stylistic details.

This was also a similar trend in the circle of Oltos. His circle is based on workshop connections, as indicated by potter signatures. The painters in his circle have low prediction values for his class; the numbers are overall low, suggesting that the neural network is very insecure concerning the attribution. This exposes the signatures as a fickle and uncertain argument, as they might have worked in the same workshops at different points of their career. However, we must also bear in mind that the model generally had more trouble recognising the Brygos Painter and Oltos compared to the Berlin Painter. Overall, the predictions seem to align with stylistic similarity especially well, which allows us to analyse the Berlin Painter's circle in more depth.

The painters in his circle are related to him in various ways and we are going to analyse them chronologically. For a long time, scholars argued that **Euthymides** [Beazley 1963, 26]; [Neils 1995] and **Phintias** [Beazley 1963, 22]; [Pécasse 1992] have been the Berlin Painters "teachers". They belong to the earlier generation and to a group called the "Pioneers" [Robertson 1992b]; [Williams 1991]. Different arguments have been proposed for each possible "teacher". Beazley wrote that the Berlin Painter "issues from their group" [Beazley 1963, 196], but also that the "subtle flow of his lines and his fine sense of composition place him nearer Euthymides" [Beazley 1918, 38]. This last judgement reflects in our model as well: Regardless of who his "teacher" may have been, our model's prediction places him closer to Euthymides, both in linework and composition.

The **Kleophrades painter** [Beazley 1933]; [Beazley 1963, 181]; [Williams 2017, 153–156], and **Myson** [Beazley 1963, 237]; [Williams 2017, 156–157] are also associated with the Pioneers' circle. Kleophrades is considered a "rival" to the Berlin Painter [Williams 2017, 153] and Myson is supposedly a "student" of Phintias. While the Kleophrades painter's prediction value is high in his full paintings, the number is lower for his actual linework, which reflects him supplying a similar market, but in a different individual style. Myson's prediction values are lower for the full paintings, but higher for the linework, maybe reflecting their common grounds regarding their "teacher". We have to be careful to draw too many conclusions from the numbers however, as they merely express our models "next best guess".

(a) Full Image						(b) Segment Set					
	Berlin	Brygos	Douris	Makron	Oltos		Berlin	Brygos	Douris	Makron	Oltos
Berlin P.											
Achilles P.	68.3	18.8	11.0	0.0	1.8		61.6	6.1	6.7	22.0	3.7
Euthymides	75.0	2.5	17.5	0.0	5.0		72.5	7.5	5.0	7.5	7.5
Hermonax	60.8	5.2	13.4	3.1	17.5		36.1	29.9	2.1	27.8	4.1
Kleophrades P.	61.0	15.7	14.8	5.6	3.0		43.3	15.7	6.2	28.2	6.6
Myson	42.9	23.8	21.4	9.5	2.4		54.8	16.7	0.0	16.7	11.9
Oionokles	77.1	4.2	6.2	0.0	12.5		60.4	25.0	4.2	6.2	4.2
Persephone P.	20.0	37.1	11.4	20.0	11.4		60.0	14.3	11.4	11.4	2.9
Phiale P.	60.0	10.7	20.0	3.3	6.0		30.0	20.0	12.7	9.3	28.0
Phintias	39.5	16.3	25.6	0.0	18.6		34.9	14.0	11.6	25.6	14.0
Providence P.	69.0	2.4	13.1	1.2	14.3		75.0	15.5	0.0	7.1	2.4
Brygos P.											
Briseis P.	12.0	36.0	17.0	18.0	17.0		39.0	39.0	2.0	16.0	4.0
Dokimasia P.	1.6	29.0	17.7	21.0	30.6		25.8	35.5	3.2	30.6	4.8
Foundry P.	8.3	34.2	17.5	17.5	22.5		15.5	36.7	5.8	38.3	4.2
P. of the Paris G.	8.0	40.0	10.7	20.0	21.3		18.7	22.7	6.7	45.3	6.7
Oltos											
Epiktetos	6.4	9.1	20.9	14.5	49.1		38.2	6.4	3.6	21.8	30.0
Nikosthenes P.	6.5	12.9	38.7	9.7	32.3		25.8	16.1	6.5	38.7	12.9
Onesimos	9.3	24.8	24.8	20.9	20.2		17.1	17.8	4.7	53.5	7.0
Pistoxenos P.	21.4	25.0	14.3	25.0	14.3		21.4	21.4	17.9	21.4	17.9
Euphronios	47.27	6.36	4.54	15.45	26.36		36.6	10.0	7.2	21.8	24.5

Table 6. Painter attribution results on the painter's friends for the model trained on full images (a) and (b) on a segment sets.

More distinct conclusions can be drawn about painters with a closer connection, his supposed "students" (Figure 12): **Hermonax** [Beazley 1963, 483]; [Oakley 2007, 72–74], the **Achilles Painter** [Beazley 1963, 986]; [Oakley 1997]; [Oakley 2007, 75–78] and the **Providence painter** [Beazley 1963, 635]; [Oakley 2007, 69–72]. In the full vase classification, our model sees a strong resemblance between the Berlin Painter, and both the Achilles Painter and Hermonax. Their values are lower in the segment-set classification – for the Achilles Painter by a little, for Hermonax significantly. According to our model, the closest student is the Providence painter, who even features higher values for the annotation sets than the full image classification. This indicates, that the similarity is not only present in his composition and motifs, but also in stylistic details such as delicately rendered hands or facial proportions. A focus on such details could make the two painters seem like the same person, and only in a closer inspection the differences become apparent as well. His other supposed "students" also share similarities with the Berlin Painter, but they often have their own twist on certain renditions of similar shapes and the similarity is often limited to specific features.



Figure 12. Paintings by painters from the Berlin Painters circle, images from the Metropolitan Museum Collection, New York. First row from left to right: Providence painter (207429, 41.162.18), Achilles Painter (214079, 41.162.139), Hermonax (205498, 41.162.19). Second row: Oionokles (207552, 28.57.11), Phiale Painter (214188, 41.162.142)

The last group of test classes features only distant connections to the Berlin Painter. The Phiale Painter is considered to be a "student" of the **Achilles Painter** [Beazley 1963, 1014]; [Oakley 1980, 58–63]; [Oakley 2007, 78–81] and **Oionokles** [Beazley 1963, 646]; [Serbeti 1989]; [Oakley 2007, 72] is considered a "follower" of the Providence painter. The **Persephone** painter was described as "akin in spirit" to the Achilles Painter (Beazley 1963, 1012; Oakley 1997, 102).

82

Since the Providence painter is close to the Berlin Painter, his "follower" Oionokles also has high prediction values for the Berlin Painter's class. Conversely, the Phiale Painter has lower prediction values, but still a higher value for the full figure classification. It may be that he would be classified closer to his own teacher, the Achilles Painter, than to the Berlin Painter, from whom his teacher may have eventually deviated. Then again, his paintings are from a much later generation. Yet the Persephone painter, who was also active much later, still shows high prediction values in the segment-set classification. This variation may therefore expose a temporal bias for later periods, as the Berlin Painters class also contains more later works than the other classes.

83

We cannot draw conclusions regarding the specific relationship between the painters based on the prediction values, but the alignment of the model's results with scholarly opinion suggests that there are underlying similarities which we have yet to interpret. While we do not go as far so to argue that the scholarly interpretations are correct, we recognize that there may be distinct similarities underlying their observations, which also get picked up by our model. Our experiments suggest that the "teacher-students-relationships" are based on complex and multi-layered features. Thus, further investigations into the details of our model are necessary to characterize the similarities and re-contextualize the scholars' interpretations.

84

Overall, the prediction values indicate that our model is able to recognize very nuanced kinds of similarities between the paintings as observed by connoisseurs. This is a good foundation for further discussion about the criteria that lead scholars to assert their propinquity in the first place. Furthermore, this experiment demonstrates the importance of the entire vase painting in addition to the segment sets since motivic and compositional congruence allow the human researcher as well as the neural network to infer similarities, just as stylistic features do. Sometimes the former might be even stronger than the latter, such as in the case of Hermonax, which may perhaps reveal a human bias in turn.

85

8. Conclusion and Outlook

The previous chapter is only a 'first look' at the results, as the analysis still reveals many peculiarities and has led to exciting case studies, which go too deep to discuss in this paper. Therefore, we would like to end our paper with a characterization of the data sampling in regards to our analysis framework, which we will be presented as brief outlook as well.

86

At its core, the individual style of a painter can be described with Beazley's term "system of renderings"; a combination of certain characteristics and their variation. With this in mind, painter attribution boils down to drawing lines between the variation of one painter against the variation of the others. To operationalize this approach, we had to take several distortions of the data into account. Being aware of these transformations, controlling them, and assessing their effect, is necessary for the creation of an analysis framework. In some instances, the distortion was deliberate, in order to emphasize important details, the "renderings" in Beazley's "system". In other instances, the potential biases and the workings of computational methods did not allow for anything else. For painter attribution, potential biases are found in general image characteristics and can be mitigated through careful painter selection and through image segmentation, intentionally skewing the representation of the original paintings to emphasize the features necessary. During this process, our annotation guidelines ensured comparability between the segments and our model architecture ensured that the segments were equally and invariantly weighted as to represent the ""systems" of Beazley's "renderings".

87

This extensive effort was necessary to ensure a comparability between the model and human researchers and to tackle the core uncertainty, fuzziness and ambiguity of the traditional method. In most cases we cannot know exactly who painted a vase. But this is not what our research is about. The work of Beazley and other scholars' hints at a phenomenon that is even more fundamental and goes beyond the question of "who painted a picture?": The phenomenon of (stylistic) variation and similarity. Where do we draw the lines? Where does one painter end and where does the other begin? The methods of machine learning offer us a new approach to these questions by simulating the referential framework and creating an exploratory space, where uncertainty and fuzziness become tangible.

To analyse these phenomena, our framework builds upon our observations in the baseline experiments and even our overall experiment design. Each data test/training data split works as a simulation of a single instances' learnability. It shows how well a model can learn a painter's characteristic by the examples provided. Some perform better than others, which points us to the uncertainties of painter attribution. Additionally, the certainty of the model and the robustness of certain instances also allows us to determine the scope of our observations. Some outliers may come with remarkable values, but turn out to be instable, while other instances may not appear very remarkable, but they appear regularly. In our further research we investigate cases of both. For this purpose, we also use dimensionality reduction techniques and their visualisations such as t-SNE (Van der Maaten and Hinton 2008) and UMAP (McInnes, Healy, and Melville 2018), which adds to the perspectivity of our approach. It adds a visual component and allows us to take the overall distribution of the data into account. All in all, we were able to re-create some nuances of painter attribution faithfully, as already shown by the first look at the results. This allows for an in-depth investigation of our own model's 'connoisseurship' in comparison with human researchers.

In any case, our first look at the results already reveals key insights into connoisseurship as a method of educated association. This adds to our understanding of image comprehension as a highly complex scholarly method, where dense information is derived from the very details of a painting. By creating a specific frame of reference with our AI-experiments, we can simulate different associative frames of scholars to explore their interpretation of similarity even deeper. Another key to understanding lies in the difference between the model's performance and human scholars. In which way do the models 30% "errors" differ from the human approach? The answer might lie in what Richard Neer calls "focused seeing" [Neer 2005]. The human is able to make an educated decisions on where to look, while a CNN blindly compares everything against everything, at least within the frame we created. Our case studies showed, that in some instances an attribution is obvious for a human expert, because the artist reveals themselves in only few, but highly identical details. Cases such as these might get lost in the statistics of machine learning.

Therefore, it is obvious, that AI is not able to replace the human expert. The relationship between the computational methods and human researchers is, on the contrary, defined by a constant interaction instead of automation (see also [Bell and Offert 2021]). Human experts have invaluable domain knowledge and contextual visual experience that cannot be adequately represented in the simulation. So, on the one hand, we have to translate and communicate our research interests to the neural network. On the other hand, it "communicates" its processing of the task back to us via evaluation metrics, prediction scores and visualisations. The "researcher" and the "artificial intelligence" are deeply intertwined, and in the end, the researchers and their interests are the key actors in the beginning and in the end of the process.

Another important takeaway from our research is that it is not always possible to distinctly isolate stylistic details from other image characteristics. Sometimes stylistic similarity correlates with general shapes, the composition and the motif of a painting. Thinking beyond painter attribution, this is a central aspect of any artistic image domain. It shows, how standard approaches might not encompass the images in all their complexity. All in all, our case study of painter attribution in Classical Archaeology provides insights into the very foundation of humanities research: Uncertainty. Ambiguity. Variation. Complexity.

Notes

[1] This work was supported by the Lower Saxony Ministry for Science and Culture (MWK), Germany, and was funded through the programme "Geistes- und Kulturwissenschaften – digital". This project was a cooperation between the Institute for Digital Humanities Göttingen, Germany and the Information Systems and Machine Learning Lab (ISMLL), Stiftung Universität Hildesheim, Germany.

[2] For example, in the Corpus Vasorum Antiquorum (CVA, *Corpus Vasorum Antiquorum* 1922-2023), which is a major project cataloguing vases of museums all over the world.

[3] Pausanias for example, described and named many sculptors in his travel report on Greece [Pausanias 1973]

[4] For each vase painting we discuss we are giving first the BAPD and then the museum and the inventory number or the number in the monograph where the vase was published. The BAPD number is the identifier in the online available Beazley Archive Pottery Database (<https://www.carc.ox.ac.uk/carc/pottery>).

[5] "I would draw attention to the bounding lines of the breasts, with the curvilinear triangle at the pit of the stomach; to the omission of the off clavicle; to the line of the hither clavicle, recurving at the pit of the neck without touching the median line of the breast; to the curved line which runs down from about hem-way along the line of the clavicle, separating shoulder and breast; to the smaller arc in the middle of the deltoid; to the indication of the trapezius between neck and shoulder; to the pair of curved lines on the upper right arm; to the projection of the wrist when the position of the hand requires it; to the two brown lines on the neck, indicating the sterno-mastoid; to the marking on the body between the lower boundary of the breast and the himation; to the form of the black lines indicating the ankle; to the pair of brown lines running from each ankle up the leg; to the forward contour of left leg and knee showing through the himation; in the himation, to the peaked folds on the left upper arm, the loose fold in the region of the navel, and the triangle where the inside of the garment shows at the shoulder"[Beazley 1922].

[6] The only other study of machine learning methods on Attic vase paintings is the one conducted by Adrian J. Ryan [Ryan 2009]. His approach differs greatly from ours however, as he focused on only one painter and artificially enhanced his dataset, rendering it useless for the analysis of style.

[7] One common approach is to increase the depth of the network by adding more layers and parameters [Simonyan and Zisserman 2014]. By doing so, the receptive field, which represents the region of the input image that influences a given feature in the network, can be enlarged. This is achieved by using smaller kernels per convolution, allowing the network to capture both fine-grained and global information. In addition to depth, other components have been introduced to improve the performance of convolutional networks. For example, the inception module [Segedy et al. 2015] incorporates multiple filter sizes within a single layer, enabling the network to capture diverse feature scales effectively. Another notable advancement is the introduction of residual connections (He et al. 2016) and densely connected blocks [Huang, Liu, Van Der Maaten, and Weinberger 2017]. These techniques address the challenge of training very deep networks which are much harder to optimize and subsequent prone to becoming stuck in the optimization process.

[8] One common example and the activation function used throughout this work is the Rectified Linear Unit (ReLU) activation [Agarap 2018] which is represented by: $f(x) = \max(0, x)$.

[9] The optimization is achieved through the gradient descent method, with the equation: $\theta_n e w = \theta_n d - \alpha \cdot \nabla J \theta$; where θ represents the model's parameters, α is the learning rate which controls the step size of the iterative updates, and $\nabla J(\theta)$ is the gradient of the loss function with respect to the parameters.

[10] A good example for a dispute among connoisseurs is the highly controversial cup in the Agora Museum in Athens (202142, P24113). For the different opinions see [Robertson 1958], [Cardon 1979] and [Kurtz 1983]. In other cases scholars focusing on specific painters and their œuvre also re-attributed works by Beazley, for example John H. Oakley for the Phiale Painter (214303, Gela, Museo Archeologico N66 [Oakley 1980, Nr. 122] or Diana Buitron-Oliver for the Painter of London E 55 (205103, Florence, Museo Archeologico Etrusco V48 [Buitron-Oliver 1995, Nr. E1 117].

[11] This involves creating modified versions of the images (such as rotations, translations, or zooming) to artificially expand and diversify the training dataset. Furthermore, samples were weighted in our loss function based on their class frequency, ensuring that underrepresented classes receive more attention during the training process. Finally, dropout was applied on the learned embeddings, a regularization technique where randomly selected neurons are ignored during training, which helps in preventing over-dependence on specific neurons and promotes generalization. It is also worth noting that the model is trained for 1000 epochs, and during each epoch, the model is presented with each figure in the training data. We implemented early stopping based on hold-out validation data, which is a technique that halts training when the model's performance ceases to improve on a set aside validation dataset, preventing potential overfitting to the training data.

[12] We are given a dataset $D = (x_1, y_1) \dots (x_N, y_N)$ sampled from an unknown distribution p that consists of Npairs of an image $x \in R^H \times W \times C$, with height H, width W and C channels, and target $y \in R^L$ for a problem with L distinct classes. The channels typically follow the RGB format. Given a loss function $l : R^N \times R^N \times R^L \rightarrow R$, we wanted to learn a model

$y: R^N \times H \times W \times C \rightarrow R^N \times L$ with minimal expected loss over the data: $E_{(x, y) \sim p} l(y, \hat{y}(x))$.

[13] Consequently, this insight leads to the development of the following formalized deep-set layer $\phi(x) = g(1/I \sum_{i=1}^I f(x_i))$, where an inner function f is applied to each element of the set x , before the outer function g is applied to the aggregation of all inner embeddings. This formalization is permutation-invariant to the order of elements in x due to the summation, while representing a simple architecture for an image-set classification task.

[14] Related work and benchmarks include the following: [Liu et al. 2019]; [Afrasiyabi, Larochelle, Lalonde, and Gagné (2022)] (similarity recognition and few-shot classification); [Maron et al. 2020] and [Zhong et al. 2018] (similarity recognition and few-shot classification).

Works Cited

- Afrasiyabi, Larochelle, Lalonde, and Gagné (2022)** Afrasiyabi, A., Larochelle, H., Lalonde, J.-F., and Gagné, C. (2022) "Matching feature sets for few-shot image classification", in *Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition*, 9014–24. Los Alamitos: IEEE Computer Society.
- Agarap 2018** Agarap, A. F. (2018) "Deep learning using rectified linear units (Relu)", arXiv Preprint. Available at: <https://doi.org/10.48550/arXiv.1803.08375>.
- Anderson 1996** Anderson, J. (1996) *Giorgione: Peintre de La" brièveté poétique"*; *Catalogue raisonné*. Paris: Lagune.
- Arrington 2017** Arrington, N. (2017) "Connoisseurship, vases and Greek art and archaeology", in [Padgett (ed.) 2017], 21–39. Princeton University Art Museum.
- Baziotopoulou-Valavani 1994** Baziotopoulou-Valavani, E. (1994) "Ανασκαφές Σε Αθηναϊκά Κεραμικά Εργαστήρια Των Αρχαϊκών Και Κλασικών Χρόνων", in *The archaeology of Athens and Attica under the democracy: Proceedings of an international conference celebrating 2500 years since the birth of democracy in Greece, held at the American School of Classical Studies at Athens, December 4-6, 1992*, 37:45–54. Oxbow Monograph. Oxford: Oxbow Books.
- Beazley 1911** Beazley, J. D. (1911) "The master of the Berlin amphora", *The Journal of Hellenic Studies*, 31:276–95.
- Beazley 1918** Beazley, J. D. (1918) *Attic red-figured vases in American museums*. Cambridge: Harvard University Press.
- Beazley 1922** Beazley, J. D. (1922) "Citharoedus", *The Journal of Hellenic Studies*, 42 (1): 70–98.
- Beazley 1930** Beazley, J. D. (1930) *Der Berliner Maler*. Vol. 2. Forschungen Zur Antiken Keramik 1. Berlin: Keller.
- Beazley 1933** Beazley, J. D. (1933) *The Kleophrades painter*. Vol. 6. Forschungen Zur Antiken Keramik 1. Berlin: Keller.
- Beazley 1956** Beazley, J. D. (1956) *Attic black-figure vase-painters*. Oxford: Clarendon Press.
- Beazley 1961** Beazley, J. D. (1961) "An amphora by the Berlin painter", *Antike Kunst*, 4 (2): 49–67.
- Beazley 1963** Beazley, J. D. (1963) *Attic red-figure vase-painters*. Oxford: Clarendon Press.
- Bell and Offert 2021** Bell, P. and Offert, F. (2021) "Reflections on connoisseurship and computer vision", *Journal of Art Historiography*, 24:1–10.
- Berrar et al. 2019** Berrar, D. (2017) "Cross-Validation". In *Encyclopedia of bioinformatics and computational biology*, Shoba Ranganathan, Michael Gribskov, Kenta Nakai, und Christian Schönbach, 542–45. Amsterdam: Elsevier, 2019.
- Boardman 1975** Boardman, J. (1975) *Athenian red figure vases: The Archaic Period*. Oxford University Press.
- Boardman 2001** Boardman, J. (2001) *The history of Greek vases: Potters, painters and pictures*. London: Thames & Hudson.
- Bolmarcich and Muskett 2017** Bolmarcich, S. and Muskett, G. (2017) "Artists' signatures on Archaic Greek vases from Athens", in *Artists and Artistic Production in Ancient Greece*, 154–76. Cambridge: Cambridge University Press.
- Bruhn 1943** Bruhn, A. (1943) *Oltos and early red-figure vase painting*. Kopenhagen: Busck.
- Buitron-Oliver 1995** Buitron-Oliver, D. (1995) *Douris: A master-painter of Athenian red-figure vase*. Forschungen Zur Antiken Keramik. II Reihe, Kerameus 9. Mainz: Verlag Philipp von Zabern.
- Burg 2007** Burg, T. (2007) *Die signatur: Formen und funktionen vom mittelalter bis zum 17. Jahrhundert*. Münster: LIT Verlag.
- Bustamente Palacio 2023** Bustamente Palacio, S. (2023) "Towards interpretable models for computer vision". Dissertation, Kaiserslautern-Landau: Rheinland-Pfälzische Technische Universität Kaiserslautern-Landau.
- Bérard and Vernant 1985** Bérard, C. and Vernant J. P. (1985) *Die bilderwelt der Griechen: Schlüssel zu Einer "Fremden" Kultur*. Translated by Ursula Sturzenegger. Vol. 31. Kulturgeschichte Der Antiken Welt. München: Verlag Philipp von Zabern.
- Büttner 2017** Büttner, N. (2017) "The hands of Rubens: On copies and their reception". *Kyoto Studies in Art History* 2 (2017): 41–53.
- Cardon 1979** Cardon, C. M. (1979) "The Gorgos Cup", *American Journal of Archaeology*, 83 (2): 169–73.
- Cline and Hasaki 2019** Cline, D. and Hasaki, E. (2019) "The connected world of potters in ancient Athens: Collaborations, connoisseurship, and social network analysis", *CHS Research Bulletin*, 7. Available at: <https://research-bulletin.chs.harvard.edu/2019/03/19/connected-world-of-potters/>.
- Cook 2013 ed.** Cook, R. M. (2013 ed.) *Greek painted pottery*. London: Routledge.
- Corpus Vasorum Antiquorum** No author (1922) *Corpus vasorum antiquorum*. Available at: <https://www.cvaonline.org/cva/Home>.
- Coulson et al. 1997** Coulson, W.D.E., Oakley, J.H., and Palagia, O. (1997) "Athenian potters and painters: The conference proceedings", in Vol. 1. Oxford: Oxbow Books.
- Deng et al. 2009** Deng, J., Dong, W., Socher, R., Li, L. J., Fei-Fei, L. (2009) "Imagenet: A large-scale hierarchical image database", in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 248–55. Los Alamitos: IEEE Computer Society.
- Driscoll 2019** Driscoll, E. (2019) "Beazley's connoisseurship: Aesthetics, natural history, and artistic development", *Natural History, and Artistic Development. Metis*, 17:101–20. Available at: <https://doi.org/10.4000/books.editionsehess.13689>.
- Elmrabet et al. 2021** Elmrabet, M. A., El Makkaoui, K., and Faize, A. (2021) "Supervised machine learning. A survey", in *2021 4th International Conference on Advanced Communication Technologies and Networking (CommNet)*, 1–10. Los Alamitos: IEEE Computer Society.
- Eschbah and Schmidt 2016** Eschbach, N. and Schmidt S. (eds.) (2016) *Töpfer Maler Werkstatt: Zuschreibungen in der griechischen vasenmalerei und die organisation Antiker Keramikproduktion*. CVA Beihefte 7. München: CH Beck.
- Felicisimo et al. 2019** Felicisimo, Á. M., Polo, M. E., Durán, G., Tortosa, T., and Rodero, A. (2019) "Rollout archaeological photography for the graphic documentation

- of cultural heritage", in Wien: Museen der Stadt Wien.
- Folego et al. 2016** Folego, G., Gomes, O., and Rocha, A. (2016) "From Impressionism to Expressionism: Automatically identifying van Gogh's paintings", in *2016 IEEE International Conference on Image Processing (ICIP)*, 141–45. Los Alamitos, (CA): IEEE Computer Society.
- Franceschini 2018** Franceschini, M. (2018) *ttische Mantelfiguren. Relevanz Eines Standardisierten Motivs Der Rotfigurigen Vasenmalerei*. Zürcher Archäologische Forschungen 5. Rahden: VML Verlag Marie Leidorf.
- Frank and Frank 2020** Frank, S. J. and Frank, A. M. (2020) "Salient slices: Improved neural network training and performance with image entropy", *Neural Computation*, 32 (6): 1222–37.
- Frank and Frank 2021** Frank, S. J. and Frank, A. M. (2021) "A neural network looks at Leonardo's(?) Salvator Mundi", *Leonardo*, 54 (6): 619–24. Available at: https://doi.org/10.1162/leon_a_02004.
- Frank and Frank 2022** Frank, S. J. and Frank, A. M. (2022) "Complementing connoisseurship with artificial intelligence", *Curator: The Museum Journal*, 65 (4): 835–68.
- Ginzburg and Davin 1980** Ginzburg, C. and Davin, A. (1980) "Morelli, Freud and Sherlock Holmes: Clues and scientific method", *History Workshop Journal*, 9:5–36.
- Giudice and Giudice 2011** Giudice, G. and Giudice, E. (2011) "Una madre in fuga su una lekythos siciliana"
- Graepler 2016** Graepler, D. (2016) "Künstlerhand Und Kennerauge. Die Zuschreibung Als Archäologisches Methodenproblem". In *Eschbach - Schmidt (2016)*, edited by Norbert Eschbach and Stefan Schmidt, 7:14–24. Corpus Vasorum Antiquorum. Deutschland. Beihefte Zum Corpus Vasorum Antiquorum. München: Verlag C.H. Beck.
- Hashemmi 2019** Hashemi, M. (2019) "Enlarging smaller images before inputting into convolutional neural network: Zero-padding vs. interpolation", *Journal of Big Data* 6 (1): 1–13.
- He, Zhang, Ren, and Sun 2016** He, K., Zhang, X., Ren, S., and Sun, J. (2016) "Deep residual learning for image recognition". In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 770–78. o. O.: IEEE Computer Society. Available at: <https://doi.org/10.1109/CVPR.2016.90>.
- Hegener and Horsthemke 2013** Hegener, N. and Horsthemke F. (2013) "Künstler-Signaturen von Der Antike Bis Zur Gegenwart". In. Petersberg: Imhof.
- Hinojosa 2009** Hinojosa, L. W. (2009) *The Renaissance, English Cultural Nationalism, and Modernism, 1860–1920*. New York: Palgrave Macmillan.
- Hoffmeyer 1943** Hoffmeyer, A. B. (1943) *Oltos and early red-figure vase painting*. Kopenhagen: Busk.
- Huang, Liu, Van Der Maaten, and Weinberger 2017** Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K. Q. (2017) "Densely connected convolutional networks", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4700–4708. o. O.: IEEE Computer Society. Available at: <https://doi.org/10.1109/CVPR.2017.243>.
- Hurwit 2015** Hurwit, J.M. (2015) *Artists and signatures in Ancient Greece*. New York: Cambridge University Press.
- Islam et al. 2021** Islam, Md.A., Kowal, M., Esser, P., Jia, S., Ommer, B., Derpanis, K.G., and Bruce, N. (2021) "Shape or texture. Understanding discriminative features in CNNs", *arXiv Preprint*, 1–15. Available at: <https://doi.org/10.48550/arXiv.2101.11604>.
- Jones and Craddock 1990** Jones and Craddock (1990) *Fake? The art of deception*. Berkeley: University of California Press.
- Khan et al. 2018** Khan, S., Rahmani, H., Ali Shah, S. A., Bennamoun, M., Medioni, G., and Dickinson, S. (2018) *A guide to convolutional neural networks for computer vision*. Cham: Springer. Available at: <https://doi.org/10.1007/978-3-031-01821-3>.
- Kipke et al. 2022** Kipke, M., Brinkmeyer, L., Bagayoko, S., Schmidt-Thieme, L., and Langgner, M. (2022) "Deep level annotation for painter attribution on Greek vases utilizing object detection", in *SUMAC '22: Proceedings of the 4th ACM International Workshop on Structuring and Understanding of Multimedia heritAge Contents*, 23–31. o. O.: Association for Computing Machinery. Available at: <https://doi.org/10.1145/3552464.355568>.
- Krizhevsky et al. 2012** Krizhevsky, A., Sutskever, I., and Hinton, G.E. (2012) "ImageNet classification with deep convolutional neural networks". Edited by F. Pereira, C. J. Burges, L. Bottou, and K. Q. Weinberger. *Advances in Neural Information Processing Systems* 25:1–9.
- Kunisch 1997** Kunisch (1997) *Makron*. Vol. 10. Forschungen Zur Antiken Keramik 2. Mainz: Verlag Philipp von Zabern.
- Kurtz 1983** Kurtz, D.C. (1983) "Gorgos' Cup: An essay in connoisseurship". *The Journal of Hellenic Studies*, 103:68–86.
- Kurtz 1986** Kurtz, D.C. (1986) "Beazley and the connoisseurship of Greek vases". *Greek Vases in the J. Paul Getty Museum*, 2:237–50.
- Langlotz 1920** Langlotz, E. (1920) *Zur Zeitbestimmung Der Strengrotfigurigen Vasenmalerei Und Der Gleichzeitigen Plastik: Beiträge*. Verlag nicht ermittelbar.
- Langmead et al. 2021** Langmead, A., Nygren, C.J., Rodriguez, P., and Craig, A. (2021) "Leonardo, Morelli, and the computational mirror". *DHQ: Digital Humanities Quarterly*, 15 (1).
- Lee et al. 2019** Lee, J., Lee, Y., Kim, J., Kosiorek, A., Choi, S., Teh, Y.W. (2019) "Set transformer: A framework for attention-based permutation-invariant neural networks", in *International Conference on Machine Learning*, 3744–53. o. O.: PMLR. Available at: <https://doi.org/10.48550/arXiv.1810.00825>.
- Levy et al. 2014** Levy, E., David, O.E., and Netanyahu, N.S. (2014) "Genetic algorithms and deep learning for automatic painter classification", in *GECCO '14: Proceedings of the 2014 Annual Conference on Genetic and Evolutionary Computation*, 1143–50. New York: Association for Computing Machinery.
- Lewis 2019** Lewis, B. (2019) *The last Leonardo: The secret lives of the world's most expensive painting*. London: William Collins.
- Li et al. 2011** Li, J., Yao, L., Hendriks, E., and Wang, J.Z. (2011) "Rhythmic brushstrokes distinguish van Gogh from his contemporaries: Findings via automated brushstroke extraction", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34 (6): 1159–76.
- Li et al. 2021** Li, Z., Liu, F., Yang, W., Peng, S., and Zhou, J. (2021) "A survey of convolutional neural networks: Analysis, applications, and prospects". *IEEE Transactions on Neural Networks and Learning Systems*, 6999–7019.
- Liedtke 2004** Liedtke, W. (2004) "Rembrandt's workshop revisited". *Oud Holland*, Nr. 117 (2004): 48–73.
- Liu et al. 2019** Liu, X., Guo, Z., Li, S., Kong, L., Jia, P., You, J., and Kumar, BVK (2019) "Permutation-invariant feature restructuring for correlation-aware image set-based recognition". In *roceedings of the IEEE/CVF International Conference on Computer Vision*, 4986–96. Los Alamitos, (CA): IEEE Computer Society.
- Lüdemann 2005** Lüdemann, P. (2005) "'Pennati Comites Linqwere Cubilia Divae': Beobachtungen Zu Giorgiones' Schlafender Venus". *Marburger Jahrbuch Für Kunstwissenschaft*, 32:93–116.

- Madhu et al. 2022** Madhu, P., Villar-Corrales, A., Kostis, R., Bendschus, T., Reinhardt, C., Bell, P., Maier, A., and Christlein, V. (2022) "Enhancing human pose estimation in ancient vase paintings via perceptually-grounded style transfer learning", *Journal on Computing and Cultural Heritage*, 6 (1): 1–17. Available at: <https://doi.org/10.1145/3569089>.
- Maffre 2001** Maffre, J.J. (2001) "Le Peintre de Brygos et Son Cercle à Athènes, à Paris et à Bâle", in *Καλλιτεχνία. Μελέτες Προς Τιμήν Της Όλγας Τζάχου-Αλεξανδρή*, 133–60. Athen: Υπουργείο Πολιτισμού ICOM-Ελληνικό Τμήμα.
- Mannack 2012** Mannack, T. (2012) *Griechische Vasenmalerei*. Verlag Philipp von Zabern in Wissenschaftliche Buchgesellschaft.
- Maron et al. 2018** Maron, H., Ben-Hamu, H., Shamir, N., Lipman, Y. (2018) "Invariant and equivariant graph networks". *arXiv Preprint*, 1–14. Available at: <https://doi.org/10.48550/arXiv.1812.09902>.
- Maron et al. 2020** Maron, H., Litany, O., Chechik, G., and Fetaya, E. (2020) "On learning sets of symmetric elements". In *International Conference on Machine Learning*, 734–44. o. O.: PMLR. Available at: <https://doi.org/10.48550/arXiv.2002.08599>.
- McInnes et al. 2018** McInnes, L., Healy, J., and Melville, J. (2018) "Umap. Uniform manifold approximation and projection for dimension reduction", *arXiv Preprint*, 1–63. Available at: <https://doi.org/10.48550/arXiv.1802.03426>.
- Monaco 2000** Monaco, M. C. (2000) *Ergasteria: Impianti Artigianali Ceramici Ad Atene Ed in Attica Dal Protogeometrico Alle Soglie Dell'ellenismo*. Rom: L'Erma di Bretschneider.
- Narvaez et al. 2022** Narvaez, M., Sagredo, I., and McGowan, N. (2022) "Painting authorship and forgery detection challenges with ai image generation algorithms: Rembrandt and 17th century Dutch painters as a case study", *IJIMAI* 7, (7): 7–13.
- Neer 1997** Neer, R. (1997) "Beazley and the language of connoisseurship", *Hephaisto*, 15:7–30.
- Neer 2005** Neer, R. (2005) "Connoisseurship and the stakes of style", *Critical Inquiry*, 31 Nr. 1 (2005): 25–52. <https://doi.org/10.1086/498001>.
- Neer 2009** Neer, R. (2009) "Connoisseurship: From ethics to evidence", in *An archaeology of representations. Ancient Greek vase-painting and contemporary methodologies*, by Dimitrios Yatromanolakis, 25–49. Athen: Institut du Livre, A. Kardamitsa.
- Neils 1995** Neils, J. (1995) "The Euthymides Krater from Morgantina", *American Journal of Archaeology*, 99 (3): 427–44.
- Noble 1988** Noble, J.V. (1988) *The techniques of painted attic pottery*. 2nd ed. London: Thames & Hudson.
- O'Shea and Nash 2015** O'Shea, K. and Nash, R. (2015) "An introduction to convolutional neural networks", *arXiv:1511.08458*. Available at: <https://doi.org/10.48550/arXiv.1511.08458>.
- Oakley 1980** Oakley, J.H. (1980) *The Phiale Painter* Vol. 8. Forschungen Zur Antiken Keramik 2. Mainz: Verlag Philipp von Zabern.
- Oakley 1997** Oakley, J.H. (1997) *The Achilles Painter*. P. von Zabern.
- Oakley 1998** Oakley, J.H. (1998) "Why Study a Greek Vase-Painter? A Response to Whitley's 'Beazley as Th Eorist'", *Antiquity*, 72 (275): 209–13.
- Oakley 2007** Oakley, J.H. (2017) "Associates and followers of the Berlin Painter", in (*Padgett 2017*), 6–84. New Haven: Yale University Press.
- Oakley 2009** Oakley, J.H. (2009) "Greek vase painting", *American Journal of Archaeology*, 113:599–627. Available at: <https://doi.org/10.3764/aja.113.4.599>.
- Oakley 2014** Oakley, J. (2014) In *Athenian potters and painters III*. Oxbow books.
- Offert and Bell 2021** Offert, F. and Bell, P. (2021) "Perceptual bias and technical metapictures: Critical machine vision as a humanities challenge", *AI & SOCIETY*, 36 (4): 1133–44. Available at: <https://doi.org/10.1007/s00146-020-01058-z>.
- Osborne 2010** Osborne, R. (2010) "The art of signing in ancient Greece", *Arethusa*, 43 (2): 231–51.
- Pace 2019** Pace, A. (2019) "Il Briseis Painter e Il Brygos Circle". In *Studi Miscellanei Di Ceramografia Greca*, 91–207. Catania: Ediarich.
- Padgett (ed.) 2017** Padgett, J. M. (ed.) (2017) *The Berlin Painter and his world: Athenian vase-painting in the early fifth century bc*. New Haven, London: Princeton University Art Museum.
- Palagia and Oakley (eds.) 1997** Palagia, O. and Oakley, J.H. (eds.) (1997) *Athenian potters and painters II. Konferenzschrift Athen 1994*. Oxford.
- Pausanias 1973** Pausanias (1973) *Pausaniae Graeciae Descriptio*. dited by Maria Helena Rocha-Pereira. Bibliotheca Scriptorvm Graecorvm et Romanorvm Teubneriana. Leipzig: Teubner.
- Pevnick 2010** Pevnick, S.D. (2010) "ΣΥΡΙΣΚΟΣ ΕΓΡΦΞΕΝ: Loaded names, artistic identity, and reading an Athenian vase", *Classical Antiquity*, 29 (2): 222–53.
- Pinney 1981** Pinney, G.F. (1991) "The nonage of the Berlin Painter", *American Journal of Archaeology*, 85 (2): 145–58.
- Pinotti 2012** Pinotti, A. (2012) "Formalism and the history of style", in *Art History and Visual Studies in Europe*, edited by M. Rampley, 75–90. Leiden: Brill.
- Pécasse 1992** Pécasse, M. (1992) "Phintias, Peintre de Coupes", *Histoire de l'art*, 19 (1): 15–28.
- Pérez 2018** Pérez, D.R. (2018) "Sir John Beazley's notebooks. A new resource for the study of Athenian figure-decorated pottery", *Hesperia*, 87 (4): 743–809.
- Robertson 1950** Robertson, M. (1950) "Origins of the Berlin Painter", *The Journal of Hellenic Studies*, 70:23–34.
- Robertson 1958** Robertson, M. (1958) "The Gorgos Cup", *The Journal of Hellenic Studies*, 0:23–34.
- Robertson 1989** Robertson, M. (1989) "Beazleys use of terms", in *Beazley Addenda. Additional References to ABV, ARV2 & Paralipomena*, edited by Thomas H. Carpenter, Thomas Mannack, and Melanie Mendonça, xii–xx. Oxford: Oxford University Press.
- Robertson 1992a** Robertson, M. (1992a) *The art of vase-painting in classical Athens*. Cambridge: Cambridge University Press.
- Robertson 1992b** Robertson, M. (1992b) "The pioneers in context". In *Wehgartner (1992)*, edited by Irma Wehgartner, 132–39. Berlin.
- Rouet 2001** Rouet, P. (2001) *Approaches to the study of attic vases: Beazley and Pottier*. Oxford: Oxford Monographs on Classical Archaeology.
- Ryan 2009** Ryan, A.J. (2009) "Computer aided techniques for the attribution of attic black-figure vase-paintings using the Princeton Painter as a model", Doktorarbeit, University of Kwazulu-Natal.
- Samek et al. 2019** Samek, W., Montavon, G., Vedaldi, A., Hansen, L. K., and Müller, K.R. (2019) *Explainable ai: Interpreting, explaining and visualizing deep*

- Santos et al. 2021** Santos, I., Castro, L., Rodriguez-Fernandez, N., Torrente-Patino, A., and Carballal, A. (2021) "Artificial neural networks and deep learning in the visual arts: A review", *Neural Computing and Applications*, 33:121–57.
- Sapirstein 2013** Sapirstein, P. (2013) "Painters, potters, and the scale of the attic vase-painting industry", *American Journal of Archaeology*, 117 (4): 493–510.
- Scheibler 1995** Scheibler, I. (1995) *Griechische Töpferkunst. Herstellung, Handel Und Gebrauch Der Antiken Tongefäße*. 2nd ed. München: Verlag C.H. Beck.
- Schmidt (ed.) 2022** Schmidt, S. (2022) *Scherben Und Geschichte: Die Absolute Datierung Bemalter Griechischer Keramik*. ol. 9. Corpus Vasorum Antiquorum. Deutschland. Beihefte Zum Corpus Vasorum Antiquorum. München: Verlag der Bayerischen Akademie der Wissenschaften; Verlag C.H. Beck.
- Schmidt 2016** Schmidt, S. (2016) "Zuschreibungen in Der Griechischen Vasenmalerei Und Die Organisation Antiker Keramikproduktion: Geschichte Und Perspektiven Der Forschung", in *Eschbach - Schmidt (2016)*, 7-13. München.
- Szegedy et al. 2015** Szegedy, C., Liu, W., Jia, Y., Sermanet, P.S., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A. (2015) "Going deeper with convolutions". In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1–9. Los Alamitos: IEEE Computer Society.
- Seifert et al. 2017** Seifert, C., Aamir, A., Balagopalan, A., Jain, D., Sharma, A., Grottel, S., and Gumhold, S. (2017) "Visualizations of deep neural networks in computer vision: A survey", in *Transparent Data Mining for Big and Small Data*, 2:123–44. Studies in Big Data. Cham: Springer. Available at: <https://doi.org/10.1007/978-3-319-54024-5>.
- Selvaraju et al. 2017** Selvaraju, R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., and Batra, D. (2017) "Grad-Cam: Visual explanations from deep networks via gradient-based localization", in *Proceedings of the IEEE International Conference on Computer Vision*, 618–26. Los Alamitos, (CA): IEEE Computer Society.
- Serbeti 1989** Serbeti (1989) "The Oionokles Painter", *Boreas*, 12:17–46.
- Simonyan and Zisserman 2014** Simonyan, K. and Zisserman, A. (2014) "Very deep convolutional networks for large-scale image recognition", *arXiv Preprint*, 1–14. Available at: <https://doi.org/10.48550/arXiv.1409.1556>.
- Sparkes 2013** Sparkes, B. (2013) *The red and the black. Studies in Greek pottery*. London: Routledge.
- Talley Jr. 1989** Talley Jr, M.K. (1989) "Connoisseurship and the methodology of the Rembrandt Research Project", *International Journal of Museum Management and Curatorship*, 8 (2): 175–214.
- Tsingarida (ed.) 2009** Tsingarida, A. (ed.) (2009) *Shapes and uses of Greek vases*. Brüssel.
- Tzachou-Alexandri 1968** Tzachou-Alexandri, O. (1968) "“ Εφορεία Κλασσικών Αρχαιοτήτων Αθηνών”, *Deltion Archaeologikon*, 23 (2): 75.
- Ugnow 2014** Ugnow, L. (2014) "Giovanni Morelli and his friend Giorgione: Connoisseurship, science and irony", *Journal of Art Historiography*, 11:1–30.
- Vakkari 2001** Vakkari, J. (2001) "Giovanni Morelli's 'scientific' method of attribution and its reinterpretations from the 1960's until the 1990's", *Konsthistorisk Tidskrift*, 0 (1–2): 46–54.
- Van der Maaten and Hinton 2008** Van der Maaten, L. and Hinton, G. (2008) "Visualizing data using T-SNE", *Journal of Machine Learning Research*, 9 (11): 2579–2605.
- Venkatesan and Li 2017** Venkatesan, R. and Li, B. (2017) *Convolutional neural networks in visual computing: A concise guide*. Boca Raton: CRC Press.
- Vierneisel (ed.) 1990** Vierneisel, K. (ed.) (1990) *Kunst Der Schale, Kultur Des Trinkens*. München: Staatliche Antikensammlungen und Glyptothek.
- Wasielewski 2023** Wasielewski, A. (2023) *Computational formalism*. Vol. 47. Leonardo. Cambridge (MA).
- Wetering et al. 2015** Wetering, E. v. d., Bruyn, J., van Nes, C.J., and Pearson, M. (2015) *A corpus of Rembrandt Paintings VI: Rembrandt's paintings revisited-a complete survey*. Dordrecht: Springer.
- Williams 1991** Williams, D. (1991) "The invention of the red-figure technique and the race between vase-painting and free painting". In *Looking at Greek Vases*, by Tom Rasmussen and Nigel Spivey. Cambridge University Press.
- Williams 2017** Williams, D. (2017) "Beyond the Berlin Painter: Toward a workshop view", in [Padgett (ed.) 2017], 144–88. New Haven, London: Yale University Press.
- Wollheim 1973** Wollheim, R. (1973) *Giovanni Morelli and the origins of scientific connoisseurship*. London: Allen Lane.
- Zaheer 2017** Zaheer, M., Kottur, S., Ravanbakhsh, S., Poczos, B., Salakhutdinov, R.R., and Smola, A.J. (2017) "Deep sets", *Advances in Neural Information Processing Systems*, 30:3–11.
- Zhong et al. 2018** Zhong, T., Arandjelovic, R., and Zisserman, A. (2018) "Compact deep aggregation for set retrieval". In *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, 1–15. Cham: Springer Nature.

Recommendations

DHQ is testing out three new article recommendation methods! Please explore the links below to find articles that are related in different ways to the one you just read. We are interested in how these methods work for readers—if you would like to share feedback with us, please complete our short evaluation survey. You can also visit our documentation for these recommendation methods to learn more.

SPECTER Recommendations

Below are article recommendations generated by the SPECTER model:

1. Computer Vision and the Creation of a Database of Printers' Ornaments, 2021, Hazel Wilkinson, University of Birmingham, UK; James Briggs, Graphcore; Dirk Gorissen, Machine Learning Ltd.
2. Capturing Captions: Using AI to Identify and Analyse Image Captions in a Large Dataset of Historical Book Illustrations, 2024, Julia Thomas, School of English Communication and Philosophy, Cardiff University; Irene Testini, Special Collections and Archives, Cardiff University
3. Transforming Information Into Knowledge: How Computational Methods Reshape Art History, 2021, Sabine Lang, Interdisciplinary Center for Scientific Computing, Heidelberg Collaboratory for Image Processing, Heidelberg University; Björn Ommer, Interdisciplinary Center for Scientific Computing, Heidelberg Collaboratory for Image Processing, Heidelberg University
4. Sensitivity and Access: Unlocking the Colonial Visual Archive with Machine Learning, 2024, Jonathan Dentler, Catholic University of Paris; German Historical Institute, Washington D.C.; Lise Jaillant, Loughborough University, UK; Daniel Foliard, Université Paris Cité, LARCA (UMR 8225); Julien Schuh, Université Paris Nanterre; Maison des Sciences de l'Homme Mondes
5. Compounded Mediation: A Data Archaeology of the Newspaper Navigator Dataset, 2021, Benjamin Lee, The Library of Congress & The University of

DHQ Keyword Recommendations

Below are article recommendations generated by DHQ Keywords:

1. Uncovering Latent Metadata in the FSA-OWI Photographic Archive, 2017, Taylor Arnold, University of Richmond; Stacey Maples, Stanford University; Lauren Tilton, University of Richmond; Laura Wexler, Yale University
2. Machine Learning Techniques For Analyzing Inscriptions From Israel, 2023, Daiki Tagami, Columbia University; Michael Satlow, Brown University
3. From Disclaimer to Critique: Race and the Digital Image Archivist, 2017, Kate Holterhoff, Georgia Institute of Technology
4. Cross-codex Learning for Reliable Scribe Identification in Medieval Manuscripts, 2024, Julius Weißmann, Media and Digital Technologies, St. Pölten University of Applied Sciences; Markus Seidl, Media and Digital Technologies, St. Pölten University of Applied Sciences; Anya Dietrich, MEG Unit, Brain Imaging Center; Martin Haltrich, Library, Klosterneuburg Abbey
5. Reviving Classical Drama: virtual reality and experiential learning in a traditional classroom, 2018, Eleni Bozia, University of Florida

TF-IDF Recommendations

Below are article recommendations generated by the TF-IDF Model:

1. Leonardo, Morelli, and the Computational Mirror, 2021, Alison Langmead, University of Pittsburgh; Christopher J. Nygren, University of Pittsburgh; Paul Rodriguez, San Diego Supercomputer Center; Alan Craig, Independent Consultant and Researcher
2. Artistic Network Toolkit (ANT): Democratizing Network Modeling in Art Historical Scholarship, 2025, Alexis Culotta, Tulane University; Aron Culotta, Tulane University
3. Dendrography and Art History: a computer-assisted analysis of Cézanne's Bathers., 2019, Melinda Weinstein, Lawrence Technological University; Edward Voss, University of Iowa; David Soll, University of Iowa
4. Unveiling the Editing Process of Japanese Demons Picture Scrolls: How Digital Humanities Played a Role in Developing a New Theory in Art History, 2023, Shoji Yamada, International Research Center for Japanese Studies
5. Researching Spanish Dance in Time and Space: A GIS for La Argentina's Ballets Espagnols, 2022, Blanca Gómez Cifuentes, Universidad Complutense de Madrid; Carlos Fernández Freire, Spanish National Research Council (CSIC); Isabel del Bosque González, Spanish National Research Council (CSIC); Idoia Murga Castro, Instituto de Historia, CSIC



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.