Peer Community In Archaeology

A demonstration of the use and finetuning of existing machine learning tools for analysing large complexes of coins

Alex Brandsen *based on peer reviews by 2 anonymous reviewers*

Chrisowalandis Deligio, Karsten Tolle, David Wigg-Wolf (2024) Supporting the analysis of a large coin hoard with AI-based methods. Zenodo, ver. 4, peer-reviewed and recommended by Peer Community in Archaeology. https://doi.org/10.5281/zenodo.8301464

Submitted: 30 August 2023, Recommended: 14 May 2024

Cite this recommendation as:
Brandsen, A. (2024) A demonstration of the use and finetuning of existing machine learning tools for analysing large complexes of coins. *Peer Community in Archaeology*, 100397. 10.24072/pci.archaeo.100397

Published: 14 May 2024

The paper outlines the ClaReNet project's exploration of computer-based methods for classifying Celtic coin series, specifically focusing on a hoard from Jersey [1]. They collaborated with Jersey Heritage and numismatists, utilising a large dataset of coin images. The process involves stages such as pre-sorting, size-based sorting, class/type identification, and die studies. They employed IT methods, including object detection and unsupervised learning, followed by supervised learning for data refinement. Collaboration with numismatic experts ensured data quality. The study highlighted challenges in classifying coins, suggesting techniques like image matching alongside convolutional neural networks (CNNs). The results demonstrate the efficacy of semi-automatic processes in coin classification, emphasising the importance of human-computer collaboration for successful outcomes.

Overall, this is a good paper, showing how we as archaeologists and numismatics can use existing tools and finetune them for our purposes; without the need for huge domain specific datasets. This research and related papers show how we can more effectively deal with the increasingly bigger data we deal with, saving time on the monotonous and labour intensive tasks, leaving us more time to deal with the big picture. An important strength of the work is the provided public software repository and the dataset. The paper is well written, and a number of images illustrate the methodology as well as the objects used.

References:

[1] Deligio, C., Tolle, K., and Wigg-Wolf, D. (2024). Supporting the analysis of a large coin hoard with AI-based methods. Zenodo, 8301464, ver. 4 peer-reviewed and recommended by Peer Community in Archaeology. https://doi.org/10.5281/zenodo.8301464

Reviews

Evaluation round #1

DOI or URL of the preprint: https://doi.org/10.5281/zenodo.8301561
Version of the preprint: 1

Authors' reply, 03 April 2024

Dear reviewers,

Thank you for reviewing our paper. We really appreciate the comments and also understand that we may not have got our message across to our target audience.

We have rewritten a large part of the paper and also added important information and our message. Our paper is aimed at numismatists who are dealing with similar problems. With our paper we did not try to develop new AI models or achieve new higher scores, we wanted to show how to take tools from IT and apply them in the process of numismatists. Our pipeline shows different tasks in a numismatist's journey, from sorting by size, by class, by die, or even starting a sorting.

Thank you again.

Decision by Alex Brandsen, posted 22 January 2024, validated 22 January 2024

Great potential, but major revisions needed

Dear Authors,

your paper describes the ClaReNet project, which analysed the extensive Le Câtillon II Celtic coin hoard, employing object recognition and scale analysis to group coins by size in the absence of detailed information. Results were presented at CAA 2022 and CAA 2023, showcasing the use of unsupervised methods and evaluating findings against expert classifications. The project also involved refining expert classifications and examining specific coin classes for die variations.

I enjoyed reading your paper, and I think it's a nice addition to the ongoing discussions around this topic. However, based on the reviewers' feedback, I feel that this paper needs major revisions to be suitable for publication.

Please check all the issues raised by the reviewers and either revise the paper or address them in your response. I hope you are not disheartened by these comments, as the research seems promising and could definitely contribute to the field, if well described and analysed.

Kind regards,

Alex Brandsen.

Reviewed by anonymous reviewer 2, 19 September 2023

The ClaReNet project is dedicated to exploring computer-based methods applied to the analysis of three distinct Celtic coin series. One of these series comprises the staters attributed to the Coriosolitae, found in the Le Câtillon II hoard in Jersey in 2012.

The primary objective of the ClaReNet project was twofold. First, it aimed to support the numismatic process by leveraging machine learning-based methods. The project addressed various aspects, including pre-sorting, classifying, and recognizing different dies. A significant emphasis was placed on employing unsupervised methods to tackle the challenge of working with an unknown dataset, devoid of any information beyond the

images themselves. Secondly, the project sought to compare its results with the classification created by the Jersey Heritage team and actively engaged numismatic experts in the evaluation process.

This paper not only presents the outcomes of the ClaReNet project but also showcases the tools, visualizations, and extensions developed during the project that proved to be valuable for facilitating communication with numismatists and integrating their expert opinions into the analysis.

The paper is well written, a number of images illustrate the methodology applied as well as the objects used.

The paper does not provide a comprehensive state of the art; neither on coin image analysis nor on the AI based methodology. While the paper covers important work on coin analysis, the motivation from both a numismatists or vision based point of view is not provided.

Furthermore the novelty of the contribution needs further clarification, how does numismatic research benefit from these findings. A main part of the paper is dedicated to (un/supervised) learning, the novelty of the contribution for computer vision is not clear.

The acquisition of the coin data is not clearly described: how is the data acquired? It seems that the numismatic standards (e.g. uniform background or illumination) for image acquisition are not applied, why is that? How is the data specified, is there something unique/special with the data?

Quantitative measure or output is scares, a comparison with state of the art or an ablation study is missing. How do you define accuracy and how is it measured?

The structure of the paper needs improvement: a clear problem formulation and motivation is missing.

Why do you need object recognition for the scale or is it size estimation? Or is it scale, and you count the coins?

Minor: The concrete implementation or acknowledging institutions during the description of the methodology is not appropriate.

For the given reasons I recommend to reject the paper.

Reviewed by anonymous reviewer 1, 17 January 2024

This paper aims to employ various machine-learning methods for the detection of coin regions in given images, their classification, and die recognition. The work was based on the images of a found hoard of Celtic coins, specifically stater coins in the found hoard.

The effort of the authors to bring this work into this publication is appreciated. The machine learning methods were used effectively in the addressed tasks. An important strength of the work is the provided public software repository and the dataset. The primary drawback of the paper is its significantly low readability, demonstrating a slight confusion in presenting computer science concepts. In this manner, I think this paper requires a significant revision.

Some notes are as follows:

1) Detection of coin region and its scale:

This is formulated as a 2-class object detection problem, where classes are the ”coin” and the ”scale bar”, and such two object classes were aimed to be detected by using a neural network. It was very confusing in the beginning what is meant by ’scale recognition’ because it sounds like an automatic calculation of the size of the coin, however, what is aimed is simply detecting the black-colored bar region in a bounding box, and such object is called ’scale’. I suggest to authors to use a less confusing notation, maybe ”scale bar”? Also, it should be mentioned clearly, the variation in the scale bar appearance in the training and test set, i.e. how many different scales they have, what the annotations look like for scale recognition. I recommend presenting a block diagram of operational steps to provide a better understanding of the approach to the user. How does the detection of the scale bar yield calculation of the coin size (this should be seen in the block diagram as well)? Also, what are the individual performances obtained for coin region detection and scale bar detection?

The model was trained on 100 training images, and then the trained model was used to find the size of a wider dataset of coins after examination on a 25-image test set, but it wasn't mentioned on which dataset this trained model was examined afterward; i.e., at the beginning of the paper, it was mentioned as this paper focused on the Staters (line 49), but then with this current analysis it is being understood as it was executed on another data because there are scale differences.

2) Clustering coins

- This section starts by mentioning Staters group was selected as a dataset. In Figure 8 its connection with the previous section can be seen but from the beginning it is very unclear the connection of this analysis with the findings in the prior section is, e.g. it is not clear whether the group of coins examined in this section was detected as staters in the previous section, or it is another dataset labeled initially by the numismatists. If Figure 8 had been shown and this divide-and-conquer approach had been presented in the very prior section of the paper (i.e. even before object detection) it would have been much better to improve the readability of the paper.

- the number of clusters was chosen as k=100, however, then we see there are 25 clusters in the results. What is the reason for that?

3) Die study

- What is the name of the parameter for value 0.3 mentioned (in line 360)? I'm guessing it is a sort of threshold to branch tree roots, but this sort of information should always be given in the text when it is seen for the first time in the text, when writing a scientific paper.

- which number of clusters was chosen for the DeepCluster application for the die study?

- Please write to the caption of Figure 15, which features were used while creating this dendrogram (ORB or Method 2)?

- which distance metric was used to compute the similarity between the image features? These should be given in the text when writing a scientific paper.

I think the authors are using the wrong taxonomizing of the adopted methodologies. It is mentioned on line 63, that the proposed approach involves two steps, namely, 1) object detection, and 2) unsupervised learning, then 3) supervised learning. It is better to add coin classes as in *unsupervised learning 'of coin classes'.* Otherwise, it doesn't look correct to categorize used approaches this way, Object detection was also done by supervised learning.

- I wonder if the authors consider publishing the die study dataset and annotations, as there is a significant lack of research on this problem due to the lack of annotated datasets. This would be certainly a significant contribution.

Abstract:

From the following sentence, it is really not possible to understand what is accomplished in this work exactly: "*First, we separated the dataset into groups of coins of different sizes using object recognition combined with the scale contained in the images. The main approach was to treat the coins independently of the underlying classification and analyze how an unsupervised method could group them.*" –> How does the first operation connect to the second operation / (mention classification was conducted on the coin image dataset collected in the first operation? - and should mention also why such prior operation was needed for the second one)?