Automatic Die Studies

Clément Cornet[1], Héloïse Aumaître[2], Romaric Besançon[1],
Julien Olivier[3], Thomas Faucher[2], Hervé Le Borgne[1],

[1]Université Paris-Saclay, CEA, List, F-91120, Palaiseau, France,
[2]Centre d'Études Alexandrines (UAR 3134 CNRS/Ifao), Alexandria, Egypt,
[3]Bibliothèque nationale de France et IRAMAT-CEB (UMR 5060, CNRS-Université d'Orléans).
Correspondence: clement.cornet/herve.le-borgne@cea.fr

1 Introduction

The quantification of ancient monetary production is crucial to understand the connections between coinage, politics, and history. (Bland, 2018). This quantification primarily relies on die studies, which compare each coin in a series (with the same types) to determine the number of engraved tools (the dies) used to strike coins. Using statistical models (Carter, 1983; Esty, 2011), it is then possible to estimate the number of dies used for each series and, consequently, the volume of coin produced. However, comparing each coin to the others in a corpus is a long and tedious manual work, which grows quadratically $\mathcal{O}(N^2)$ with the size N of the corpus, limiting current dies studies to a few thousand coins.

2 Contribution

We propose a fully automated method for conducting die studies on low-resolution RGB images. Our approach involves extracting robust and fast deep learning features from coins, which are then filtered using a Random Sample Consensus method that operates without requiring hyperparameter settings. Subsequently, the matches between coins are used to construct a graph, and the coins are clustered using a graph clustering algorithm. To determine the crucial hyperparameter for graph clustering, we propose to rely on the intrinsic structural consistency of potential partitions. Hence, this method operates entirely without human intervention.

3 Related Works

While most automated coin analysis systems are focused on coin types recognition (Manzoor et al., 2022; Cooper and Arandjelović, 2020) and corpus analysis (Deligio et al., 2024), some works tackle the problem of automating die studies. The seminal contribution in this direction is CADS (Taylor, 2020). It consists in extracting keypoints descriptors for every coin using ORB (Rublee et al., 2011) and finding the corresponding matches. From these, they construct a pairwise dissimilarity measure upon the descriptors, and perform an agglomerative clustering algorithm (Kaufman and Rousseeuw, 1990) using this measure. The clusters obtained correspond to the coins struck by a same die. However, CADS is semi-automatic, as it requires human intervention at preprocessing and clustering stages. In a preprint Heinecke et al. (2021) propose an automated ("unsupervised") approach based on a similar pipeline as CADS, with Gaussian process keypoint extraction and a Bayesian distance microclustering algorithm (Natarajan et al., 2024). However, their code is only partially released, which makes it hard to reproduce their results. Our approach differs from both these works, with the use of more efficient deep learning local features and a fully automatic clustering method. The Riedones3D approach (Horache et al., 2021) exhibits promising results for automated die studies, relying on 3D-scanning every coin of the corpus. It is nevertheless limited in practice since it requires to re-scan every corpus with 3D-scan.

4 Proposed Approach

Similarly to the way numismatists compare coins in die studies, we look for matching features from every pair of coins. Differing from previous works (Taylor, 2020; Heinecke et al., 2021), we use deep-learning matches, namely XFeat (Potje et al., 2024) that obtains remarkable results on image matching benchmarks (Mishkin et al., 2015; Li and Snavely, 2018), while being extremely fast, yet suitable for larger-scale die-studies.

We use MAGSAC++ (Barath et al., 2020) for filtering inlier matches from outliers. The number of filtered matches between two coins can then be interpreted and handled as a similarity measure.

Then, repeating the process for every pair of coins in a collection produces a pairwise similarity matrix M. We derive a pairwise dissimilarity matrix as $D = \max(M) - M$, better suited for further analysis. We construct a graph $\mathcal{G}$ which nodes correspond to each coin of the considered corpus. Two nodes are connected if the number of matches between corresponding coins exceeds a threshold τ. When τ is chosen appropriately, coins should be connected only if they were struck by the same die. This approach of using an unweighted graph aligns more closely with traditional die studies.

To obtain our final partition, we perform graph clustering over $\mathcal{G}$, using the Label Propagation Clustering algorithm (Raghavan et al., 2007). This approach first assigns a unique label to each node. It then iteratively updates each node's label to the most common label among its neighbors, processing nodes in a random order until convergence.

To determine the optimal threshold τ^*, we compute the partitions corresponding to every possible value of τ. For each of them, we compute the Silhouette Coefficient (Rousseeuw, 1987) that does not require ground truth clustering and only evaluates the coherence of the inner structure of a partition, given a dissimilarity measure. For a sample x_i that is in the k^{th} cluster C_k, it assesses whether it is closer to the other samples of its own cluster, or to the next-nearest cluster. Finally, we keep the partition that has the highest Silhouette Coefficient. The resulting Adaptive Graph Label Propagation (AGLP) approach thus allows to determine the groups that correspond to each die without any manual hyperparameter setting.

5 Experimental Results

We propose a first baseline $CADS - AG^*$ that reproduces the semi-automatic approach CADS with a human user able to find the best possible cut for the automatic clustering. The second baseline is actually automatic, using a HDBSCAN clustering from the pairwise distance matrix, with a minimum cluster size of 2 ($CADS - HDBSCAN$).

We perform the evaluation on two collections of coins, with their associated die analysis : the Paphos (Olivier and Faucher, 2017) and Tanis 1986 (Faucher et al., 2017) hoards, with respectively 2484 and 295 coins. We consider the coin images at a resolution of 288 × 288 pixels for Paphos, and 480 × 512 for Tanis 1986.

To evaluate the results, we use three standard external clustering validity measures, namely the Adjusted Rand Index (ARI) (Vinh et al., 2009), the Adjusted Mutual Information (AMI) (Meilă, 2007) and the Fowlkes-Mallows Index (FMI) (Fowlkes and Mallows, 1983).

The experimental results (Table 1 and Appendix B) exhibits signantly higher performances than the baselines. Thanks to this automated approach, which code is publicly released, ancient historians will be able to carry out die studies on a much larger scale, and thus deepen our understanding of the ancient world. This will be particularly useful in the case of political entities for which die studies could not be completely carried out due to the size of the available corpora, such as Athen, Aegina or Roman Empire. Free of historical bias and able to deal with unprecedented quantities of material, from several different mints, this should reveal links that humans would have thought impossible or unlikely a priori

However, the two corpora considered in this paper can be considered as quite homogeneous, since all the coins are represented by RGB pictures with about the same size, illumination conditions, viewpoint and resolution. Other collections likely to be the subjects of die studies can be much more heterogeneous when the various coins have been captured by different historians (or individuals), at different times, with different photo equipment. In some cases, all we have is a scan of the negative of the original photo, or the one printed in a historical research article, or even the image of a (plaster) cast of the coins. Therefore, the challenge will be to apply the proposed approach to these cases, in particular because the keypoints detector currently used is adapted to natural RGB images. Matching such heterogeneous data is an ambitious challenge, but is necessary to address die studies in all their diversity.

This work was published in (Cornet et al., 2024).

Approach	AMI	ARI	FMI
$CADS - HDBSCAN$	0.747	0.678	0.703
$CADS - AG^*$	0.664	0.483	0.517
Ours	**0.981**	**0.978**	**0.978**
$CADS - HDBSCAN$	0.532	0.437	0.471
$CADS - AG^*$	0.499	0.382	0.445
Ours	**0.920**	**0.908**	**0.909**

Table 1: Die study results on the Paphos (top) and Tanis 1986 (bottom) collections

References

Daniel Barath, Jana Noskova, Maksym Ivashechkin, and Jiri Matas. 2020. Magsac++, a fast, reliable and accurate robust estimator. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1304–1312.

R.F. Bland. 2018. Quantifying the size of a coinage: die studies or coin finds. In N.T. Elkins and J. DeRose Evans, editors, *Concordia Disciplinarum. Essays on Ancient Coinage, History, and Archaeology in Honor of William E. Metcalf*, pages 223–234. Numismatic Studies 38, New York.

Giles F. Carter. 1983. A simplified method for calculating original number of dies from die link statistics. *Museum Notes (American Numismatic Society)*, 28:195–206.

Jessica Cooper and Ognjen Arandjelović. 2020. Learning to describe: a new approach to computer vision based ancient coin analysis. *Sci*, 2(2):27.

Clément Cornet, Héloïse Aumaître, Romaric Besançon, Julien Olivier, Thomas Faucher, and Hervé Le Borgne. 2024. Automatic die studies for ancient numismatics. In 3^{rd} *International Workshop on Artificial Intelligence for Digital Humanities (AI4DH@ECCV)*.

Chrisowalandis Deligio, Karsten Tolle, and David Wigg-Wolf. 2024. Supporting the analysis of a large coin hoard with ai-based methods. *Zenodo preprint*.

W. Esty. 2011. The geometric model for estimating the number of dies. In François de Callataÿ, editor, *Quantifying monetary supplies in Greco-Roman times*, pages 43–58. Pragmateiai 19, Bari.

Thomas Faucher, Julien Olivier, Philippe Brissaud, and Christelle Desbordes. 2017. EH 208. Trésor de Tanis, 1986. In Thomas Faucher, Andrew Meadows, and Catharine Lorber, editors, *Egyptian Hoards I, The Ptolemies*, number 168 in Bibliothèque d'étude, pages 203–222.

Edward B Fowlkes and Colin L Mallows. 1983. A method for comparing two hierarchical clusterings. *Journal of the American statistical association*, 78(383):553–569.

Andreas Heinecke, Emanuel Mayer, Abhinav Natarajan, and Yoonju Jung. 2021. Unsupervised statistical learning for die analysis in ancient numismatics. *arXiv preprint arXiv:2112.00290*.

Sofiane Horache, Jean-Emmanuel Deschaud, Francois Goulette, Katherine Gruel, Thierry Lejars, and Olivier Masson. 2021. Riedones3d: a celtic coin dataset for registration and fine-grained clustering. *arXiv preprint arXiv:2109.15033*.

L Kaufman and PJ Rousseeuw. 1990. Agglomerative nesting (program agnes) in finding groups in data: an introduction to cluster analysis. *Hoboken*.

Zhengqi Li and Noah Snavely. 2018. Megadepth: Learning single-view depth prediction from internet photos. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2041–2050.

Sehrish Manzoor, Nouman Ali, Muhammad Raees, Khan Awais Khan, Muhammad Usama Ayub, and Afzal Ahmed. 2022. Ancient coin classification based on recent trends of deep learning. In *VIPERC*.

Marina Meilă. 2007. Comparing clusterings—an information based distance. *Journal of multivariate analysis*, 98(5):873–895.

Dmytro Mishkin, Jiri Matas, Michal Perdoch, and Karel Lenc. 2015. Wxbs: Wide baseline stereo generalizations. *arXiv preprint arXiv:1504.06603*.

Abhinav Natarajan, Maria De Iorio, Andreas Heinecke, Emanuel Mayer, and Simon Glenn. 2024. Cohesion and repulsion in bayesian distance clustering. *Journal of the American Statistical Association*, 119(546):1374–1384.

Julien Olivier and Thomas Faucher. 2017. *Egyptian Hoards I: The Ptolemies*, chapter EH 69. Le trésor de Paphos (IGCH 147 ; CH 2.106 ; 4.68). Institut Français d'Archéologie orientale.

Guilherme Potje, Felipe Cadar, André Araujo, Renato Martins, and Erickson R Nascimento. 2024. Xfeat: Accelerated features for lightweight image matching. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2682–2691.

Usha Nandini Raghavan, Réka Albert, and Soundar Kumara. 2007. Near linear time algorithm to detect community structures in large-scale networks. *Physical review E*, 76(3):036106.

Simone Romano, Nguyen Xuan Vinh, James Bailey, and Karin Verspoor. 2016. Adjusting for chance clustering comparison measures. *Journal of Machine Learning Research*, 17(134):1–32.

Peter J Rousseeuw. 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 20:53–65.

Ethan Rublee, Vincent Rabaud, Kurt Konolige, and Gary Bradski. 2011. Orb: An efficient alternative to sift or surf. In *2011 International conference on computer vision*, pages 2564–2571. Ieee.

Zachary McCord Taylor. 2020. The computer-aided die study (cads): A tool for conducting numismatic die studies with computer vision and hierarchical clustering. Computer Science Honors Theses. 54.

Nguyen Xuan Vinh, Julien Epps, and James Bailey. 2009. Information theoretic measures for clusterings comparison: is a correction for chance necessary? In *Proceedings of the 26th annual international conference on machine learning*, pages 1073–1080.

A Illustration

In Figure 1 we illustrate a air of coins that have been struck with the same die (numbered 24) and a coin of the same type struck with another die.

B Details on Quantitative Results

Baseline: a fully automatic CADS approach

In order to compare our approach to previous works, we reproduce the CADS model (Taylor, 2020), with some adaptations to make the model fully automated, since the original model requires manual operations:

- CADS only considers matches within a user-defined circle, in order to filter out edges. Human intervention is needed because the optimal radius of the circle varies from one coin to another. Since this process is not suited for automation, we ignore this filtering step in our fully automatic version of the model;
- CADS builds a hierarchy between coins with agglomerative nesting (Kaufman and Rousseeuw, 1990), but does not automatically decide where to cut it to build a partition, leaving the decision to the user.

We propose a first baseline $CADS - AG^*$ that reproduces the semi-automatic approach CADS with a human user able to find the best possible cut for the automatic clustering. The second baseline is actually automatic, using a HDBSCAN clustering from the pairwise distance matrix, with a minimum cluster size of 2 ($CADS - HDBSCAN$).

Datasets

We perform the evaluation on two hoards of coins, with their associated die analysis: the Paphos (Olivier and Faucher, 2017) and Tanis 1986 (Faucher et al., 2017) hoards, with respectively 2484 and 295 coins. We consider the coin images at a resolution of 288 × 288 pixels for Paphos, and 480 × 512 for Tanis 1986. In (Taylor, 2020), the CADS model was evaluated on a subset of 200 coins of the Paphos hoard.

Since only a small proportion of coins from Antiquity have been found to this day, coins follow uneven distributions across dies on both collections (Figure 2). For example, 249 out of 2484 coins from the Paphos collection (9.98%) are singletons - the only ones to correspond to their respective die. Also note that we only consider the obverse of each coin (not the reverse). For manufacturing reasons (mostly for Greek coins), the reverse die used to break more often, leading to 1856 (74.7%) singletons on the Paphos hoards.

Metrics

To evaluate the results, we use three standard external clustering validity measures, that evaluate the similarity between two partitions U and V. They can then be used to estimate the quality of a computed partition by comparing it to a ground truth. All those metrics range from zero to one and are maximized for similar partitions.

The Adjusted Rand Index (ARI) (Vinh et al., 2009) is based on the Rand Index (RI) that evaluates the proportion of sample pairs that are correctly partitioned, adjusted for chance against unbalanced clusters.

$$ARI = \frac{RI - E[RI]}{\max(RI) - E[RI]} \tag{1}$$

Similarly, the Adjusted Mutual Information (AMI) (Meilă, 2007) is an adjusted version of the Mutual Information (MI) in an Information Theory sense. In the general case, AMI and ARI are often close, but AMI is supposed to be more relevant when the ground truth is unbalanced (Romano et al., 2016). Noting $H(.)$ the Shannon Entropy, the AMI of two partitions U and V is defined by :

$$AMI(U, V) = \frac{MI - E[MI]}{mean(H(U), H(V)) - E[MI]} \tag{2}$$

Lastly, the Fowlkes-Mallows Index (FMI) (Fowlkes and Mallows, 1983) is the geometric mean of pairwise Precision and Recall, and is generally more relevant with noisy data or unrelated partitions. We also report pairwise Precision and Recall, since a significant difference between both could help explaining poor results in some cases.

Quantitative Results

We present a comparison of our model with the fully automatic CADS model (Taylor, 2020) on the datasets Paphos (Table 2) and Tanis (Table 3). We first observe that, in the CADS approach, the HDBSCAN clustering performs better than the best obtainable result with the agglomerative nesting algorithm that was used in (Taylor, 2020). We also note that our method clearly outperforms CADS over every evaluation measure on both datasets, with most metrics above 0.90.

(a) Die 24

(b) Die 24

(c) Die 52

Figure 1: Differences between coins from different dies, Paphos (Olivier and Faucher, 2017). Coins 27 (1a) and 28 (1b) were struck by the same die, while coin 95 (1c) was struck from a different one.

Approach	AMI	ARI	FMI	Precision	Recall
$CADS - HDBSCAN$	0.747	0.678	0.703	0.914	0.540
$CADS - AG^*$	0.664	0.483	0.517	0.742	0.360
Ours	**0.981**	**0.978**	**0.978**	**0.974**	**0.981**

Table 2: Die study results on the Paphos collection

Approach	AMI	ARI	FMI	Precision	Recall
$CADS - HDBSCAN$	0.532	0.437	0.471	0.674	0.329
$CADS - AG^*$	0.499	0.382	0.445	0.769	0.258
Ours	**0.920**	**0.908**	**0.909**	**0.932**	**0.888**

Table 3: Die study results on the Tanis 1986 collection

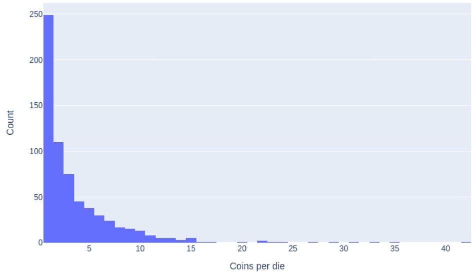

Figure 2: Distribution of number of coins per die, Paphos

While the performances are much higher than CADS, some differences remain in comparison to die studies conducted by humans. The fraction of dies that are correctly identified among all those that should have been identified is reflected by the recall. The fraction of dies correctly identified among all those that have been identified is reflected by the precision. Since the number of dies is usually quite small in practice, a manual inspection of the identified groups can be conducted, leading to improved precision. However, the errors in terms of recall remain problematic since fixing them requires to inspect the full collection, thus losing the benefits of automation.