AIJRRLSJM VOLUME 8, ISSUE 3 (2023, MAR) (ISSN-2455-6602)ONLINE
Anveshana's International Journal of Research in Regional Studies, Law, Social Sciences, Journalism and Management Practices

STUDY ON INTERPRETATION OF IMAGES ON ANCIENT COINS

Neeraj Kumar
Research Scholar
Department of History
OPJS University, Churu, Rajasthan.
nesuneeraj1985@gmail.com

Dr. Nisha sharma
Research Guide
Department of History
OPJS University, Churu, Rajasthan.

Abstract

A variety of issues that fall under the broad category of autonomous, computer vision-based study of ancient coins have gained more and more attention in recent years. Despite this study endeavor, the outcomes obtained by the state of the art are still subpar and far from performing well enough for any practical application. We provide a number of contributions in the current article that we hope will be helpful to the interested community. First, we clarify that the approach of visual coin matching, which is used in every paper on the subject currently in print, is not of practical interest because there are far more types of ancient coins than there are of those types that have been imaged, whether in digital form or otherwise. Instead, we contend that attention should be paid on comprehending the semantic information contained in coinage. Therefore, we provide a unique approach that extracts and associates semantic ideas with the appropriate coin images using real-world multimodal input, and then use a novel convolutional neural network to learn the appearance of these concepts. We provide very encouraging findings using empirical data on the biggest data collection of historical coins in the actual world.

Introduction

The study of money, including coins, bills, and tokens, is known as numismatics. Scholars, amateur collectors, and professional dealers all find this topic to be quite interesting since it offers significant cultural and historical insights. Theft detection, finding and classifying findings, and forgery prevention are all significant applications of machine learning in this area. Ancient coins are the subject of the current study.

Due to centering, wear, patination, and variations in how the same semantic components are shown artistically, ancient coins of the same kind might seem quite differently from one another. This presents a variety of technological difficulties and makes it challenging to accurately determine which ideas are represented on a specific coin using autonomous machine learning and computer vision approaches. Due to the specialized expertise needed, it is common practice for experienced dealers or researchers to identify and classify ancient coins. This is a time-consuming procedure that requires years of skill. A coin's value is determined by its denomination, the monarch it was struck under, the time and location of its minting, and human experts. There are several distinct traits that are often employed for attribution, including:

- Physical characteristics such as weight, diameter, die alignment and colour,
- Obverse legend, which includes the name of the issuer, titles or other designations,
- Obverse motif, including the type of depiction (head or bust, conjunctional or facing, etc.), head adornments (bare, laureate, diademed, radiate, helmet, etc.), clothing (draperies, breastplates, robes, armour, etc.), and miscellaneous acces- sories (spears, shields, globes, etc.),
- Reverse legend, often related to the reverse motif,
- Type and location of any mint markings.

1.1 Relevant prior work

Because the majority of currently used algorithmic techniques to coin recognition rely on local features, they perform poorly because spatial linkages between components are lost. Approaches that split a coin into segments have been suggested to get over this constraint, however they make the assumption that coins are properly centered, precisely recorded, and almost circular in form. Recent research demonstrates that the brittleness of these assumptions causes existing techniques to perform badly on real-world data. Despite the fact that the legend may be a useful source of information, it can be difficult to identify coins based only on their legend since it is often adversely impacted by wear, lighting, and minting faults. Last but not least, various other difficulties arise during and as a result of the automated data preparation process, including segmentation, scale, orientation, and color normalization.

The use of visual matching and the presumption that the unknown query coin is one of a small number of gallery kinds restrict all previous research in this field, which is especially essential to mention in the context of the current study. This notion, however, is unfounded since there are thousands of distinct coin kinds. Therefore, the purpose of this project is to investigate the feasibility of comprehending the creative material displayed on a coin. This knowledge could then be used to text-based matching, opening up access to a far wider range of coin kinds without the need for them to be recorded in the training data. Our ultimate goal is to discover the distinct semantic aspects that are represented on a coin in order to try to explain it in a manner similar to that of a human expert.

Problem specification, constraints, and context

Many different picture interpretation tasks have had success using deep learning. Recent groundbreaking research on its use to coin pictures has shown very promising results, beating more conventional methods by an order of magnitude. The fact that our data set is weakly supervised and that samples are only labeled at the image level by a corre- sponding largely unstructured textual description, despite the fact that the actual semantic elements of interest themselves occupy a relatively small area of the image, represents a significant difference in the nature of the challenge we address here. Compared to fully trained data (pixel level labels), this kind of data is far more common and simpler to get, but it also presents a much bigger problem.

Challenge of weak supervision

Remember that our main objective is to determine whether or not the reverse of an unidentified question coin contains several significant semantic features. In addition, we want a scalable method that can be applied to a large list of components that have been automatically retrieved. We selected to utilize photos that are comparatively clean since this is the first attempt at completing the specified goal; doing so helped us to isolate the difficulties of localizing the coin in an image and segmenting it off. Fig. 1 shows some examples of the photographs utilized.

Fig. 1 Typical images used in the present work

Each coin picture is linked to an unstructured text description of the coin supplied by the expert dealer selling the coin in order to aid in the automated extraction of significant semantic features and the learning of their visual appearance. This textual meta-data contains a wide range of information, such as catalogue references, minting dates and locations, provenance, etc. It always includes descriptions of the coin's obverse and reverse as well as the issuing authority (usually the emperor on the obverse).

The practical need for a scalable technique has previously been emphasized. It is therefore only possible to say, at best, whether or not a specific semantic element (as inferred from text; see next section) is present in an image due to the nature of the available data. Finer labeling, i.e., describing the location or shape of the element, is not possible due to the amount of human labor required (there are far too many potential elements and the amount of data that would require labeling is excessive), necessitating weak supervision of the kind.

Preparing and cleaning up data

Unstructured, weakly standardized, and heterogeneous data, as in most real-world applications, provide a significant difficulty in the wide scope of interest in the current study due to potential incorrect labeling, peculiarities, etc. Therefore, thorough pre-processing and data cleanup are essential steps for promoting learning.

the pre-processing of images All photos are first cropped to a square bounding box that only includes the matching coin's reverse, and the resulting image is then isotropically resized to a consistent scale of 300 300 pixels. The bust of an emperor is often shown on the obverse of a coin, while the semantically significant portions are found on the reverse.

Semantic extraction from text We utilize the unstructured text files to analyze the frequency of word occurrences and, in particular, concentrate on the most prevalent ideas in order to choose which semantic features to focus on (see Figs 3 and 4). The aforementioned choice was chosen to guarantee that there is enough training data since our study is the first to try to solve the issue at hand.

Fig. 2 Examples of depictions of reverse motif elements the present work focuses on

Before labeling the data, we first eliminate redundant words and punctuation from the attribution text files. We utilize a translation API (googletrans) to produce translations of each term since the data included attributions in French, Spanish, German, and English in addition to the other four languages. As a result, to make greater use of the data that is available, while creating the data set used to understand the idea of "horse," photos with related text that includes the terms "horse," "caballo," "cheval," and "pferd" are also included as positive examples. Included are the plurals, synonyms, and other terms that are closely related to the subject matter as well as their translations.

stratification and randomization For each of the chosen components we shuffle the samples before constructing training, validation, and test sets, with a ratio of 70% training set, 15% validation set, and 15% test set. We employ stratified sampling to assure equal class representation in order to correct the under-representation of positive cases.

Errors in data

Even after preprocessing, the final data in the context of the current issue, as in most large-scale, real-world applications of computerized data analysis, will include a certain percentage of inaccurate data. The types of errors vary from improper label assignments to improperly produced photos.

As an example, we found many cases where the keyword used to indicate whether or not a certain coin picture includes the desired element was really referring to the coin's obverse rather than the reverse, which is what we are interested in. Shields are the most often affected by this, which causes inaccurate labeling when creating the associated data sets. Our assumption later supported by empirical evidence is that incorrect labeling of this kind is not systematic in nature. As a result, the coherence both visual and non-visual of data that has been correctly labeled will outweigh the cumulative impact of a relatively small number of incorrect labels.

We also discovered that several of the original photos had odd layouts for the obverse and reverse of the respective coins, which led to distortions during preprocessing. These would be considered negative samples and, since they seem to be virtually solely of the auction listing kind mentioned above, cumulatively rather inconsequential.

Proposed framework

Using a unique deep convolutional neural network that is loosely based on AlexNet, we explain it in this research as a way to learn the appearance of semantic components and determine whether or not an unknown question coin includes the stated element. We are able to locate the element when it is present by creating heatmaps using the occlusion approach, even if we do not utilize this information in the current work for anything other than the analysis of findings.

Model topology

Our network has five convolutional layers in total (three of which are followed by a max pooling layer), which are described in Fig. 6 and Table 1. The output is then flattened and passed through three fully connected layers. High level characteristics in the input data are found by the convolutional and pooling layers, which are then used by the fully connected layers to provide a class prediction.

The convolutional layers make it possible to recognize characteristics in the data regardless of their placement. A number of hyperparameters that we define for each convolutional layer affect the output of the layer. These are the stride, depth (also known as the number of kernels), kernel size, and padding. It is required to use many kernels in order to understand a variety of characteristics. The complexity of the features we are able to detect is dependent on the number of convolutional layers present in a network. For instance, one can reasonably expect to detect only simple, low-level features such as edges from the original 'raw' pixel inputs in the first layer, and then from those edges learn simple shapes in the second layer, and then more complex, higher level features from those simple shapes, and so on.

The pooling layers may be compared to downsampling since they try to minimize output size while preserving each kernel's most crucial data. By lowering the amount of parameters we need to learn, pooling also reduces the network's complexity. Additionally, it helps the network become less sensitive to minute changes, translations, and distortions in the input picture. Utilizing max-pooling, each kernel's highest value contributes one element to the output of the pooling layer in our approach.

The dense layers, which are also known as fully-connected layers, convert the high level features discovered by the earlier layers into a class prediction. The thick layers in our model use dropout, a regularization method that combats overfitting by randomly eliminating nodes at each iteration, to prevent overfitting. A hyperparameter, which we set to 0.5 in accordance with earlier work on CNNs, determines how many nodes are deleted.

The greatest number of epochs we empirically determined to be more than adequate is 200, and we utilize a batch size of 24, with training always coming to a close far before that point. We determine at each epoch if the loss is less than 0.001 or whether the loss has not improved for 30 consecutive epochs. Training is stopped if one of the prerequisites is met. This is done in part to prevent models from overfitting and in part to save time in cases when a model or particular collection of hyperparameters are obviously underperforming.

Adaptive moment estimation (Adam) optimization, a kind of gradient descent, is used to train the model, with the learning rate and momentum being calculated for each weight separately. We utilize cross entropy loss since we are conducting categorization. To overcome issues caused by vanishing gradients, the rectified linear unit is used at the activation function.

Results and discussion

Our method produces a high degree of accuracy across training, validation, and test data. The fact that the test and validation scores are not noticeably lower than the training scores shows that there is not much indication of overfitting and points to a well-designed architecture and a large data set. The models used to identify cornucopiae and paterae are particularly successful, which is probably because there is not much variation in artistic representations of these elements. In contrast to shields, horses, and eagles, which are depicted in a variety of orientations, positions, and styles.

Learnt salient regions

The occlusion approach is used to rate the significance of several picture regions in relation to the job at hand. In a nutshell, the procedure entails artificial occlusion by a uniform kernel and measurement of the classification performance difference between inputs that are not occluded and those that are, with a bigger difference indicating a higher relevance. Previous research on ancient coin analysis using computer vision has shown how beneficial this method is for interpreting empirical findings. In this study, we adopt the usage of three kernel sizes, 32 32, 48 48, and 64 64 pixels, to provide resistance to relative size.

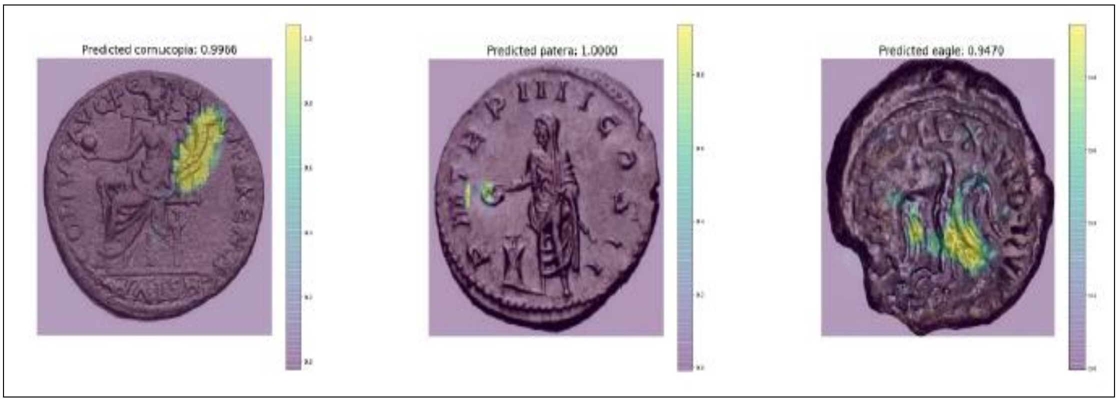

Fig. 3 Examples of automatically identified salient regions corresponding to a cornucopia, a patera, and a shield, respectively

Figure 8 provides typical examples of detected salient areas for various semantic components. It is clear that our algorithm successfully recognizes the most distinctive sections of the components, such as the elongated paterae, twisting patterns of cornucopiae, and eagle feather patterns. After doing so, we came to the reassuring conclusion that shields are inherently harder to identify because they have the least amount of distinctive detail, the greatest variability in appearance (shown from the front, they appear circular, while shown from the side, they assume an arc-like shape), and the style of depiction.

Conclusions

We achieved some significant advancement in the area of computer-based analysis of ancient coins in this work. First, after outlining the practical shortcomings of visual matching of ancient coin pictures, we provide the first argument against its usage. Instead, we made the case that focus should be placed on the semantic interpretation of coin pictures and presented the results of the first attempt at this difficult endeavor. In particular, we described a novel method that combines unstructured text analysis with visual learning through the use of a convolutional neural network. This method establishes shaky connections between semantic elements discovered on ancient coins and the corresponding images in order to learn the appearance of the aforementioned elements. We conducted the biggest experiment in the body of literature to show the effectiveness of the suggested technique utilizing photographs of coins taken from 100,000 auction lots. We displayed the visualization of learned concepts on particular examples of coins in addition to a thorough statistical analysis, demonstrating that our system is actually producing the right correlations. We are certain that our contributions will help to guide future work and provide new, intriguing research opportunities.

References

1. Agarap, A.F.: Deep learning using rectified linear units (ReLU). arXiv p. 1803.08375 (2018)

2. Anwar, H., Zambanini, S., Kampel, M.: Supporting ancient coin classification by image- based reverse side symbol recognition. In Proc. International Conference on Computer Analysis of Images and Patterns pp. 17–25 (2013)

3. Anwar, H., Zambanini, S., Kampel, M.: Coarse-grained ancient coin classification using image-based reverse side motif recognition. Machine Vision and Applications 26(2), 295– 304 (2015)

4. Arandjelovi´c, O.: Automatic attribution of ancient Roman imperial coins. In Proc. IEEE Conference on Computer Vision and Pattern Recognition pp. 1728–1734 (2010)

5. Arandjelovi´c, O.: Reading ancient coins: automatically identifying denarii using obverse legend seeded retrieval. In Proc. European Conference on Computer Vision 4, 317–330 (2012)

6. Conn, B., Arandjelovi´c, O.: Towards computer vision based ancient coin recognition in the wild – automatic reliable image preprocessing and normalization. In Proc. IEEE In- ternational Joint Conference on Neural Networks pp. 1457–1464 (2017)

7. Fare, C., Arandjelovi´c, O.: Ancient Roman coin retrieval: a new dataset and a systematic examination of the effects of coin grade. In Proc. European Conference on Information Retrieval pp. 410–423 (2017)

8. Janocha, K., Czarnecki, W.M.: On loss functions for deep neural networks in classification. arXiv p. 1702.05659 (2017)

9. Kampel, M., Zaharieva, M.: Recognizing ancient coins based on local features. In Proc. International Symposium on Visual Computing 1, 11–22 (2008)

10. Kinga, D., Adam, J.B.: A method for stochastic optimization. In Proc. International Conference on Learning Representations 5 (2015)

11. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolu- tional neural networks. Advances in Neural Information Processing Systems pp. 1097–1105 (2012)

12. Mattingly, H.: The Roman Imperial coinage., vol. 7. Spink (1966)

13. Rieutort-Louis, W., Arandjelovi´c, O.: Bo(V)W models for object recognition from video. In Proc. International Conference on Systems, Signals and Image Processing pp. 89–92 (2015)

14. Rieutort-Louis, W., Arandjelovi´c, O.: Description transition tables for object retrieval using unconstrained cluttered video acquired using a consumer level handheld mobile device. In Proc. IEEE International Joint Conference on Neural Networks pp. 3030–3037 (2016)

15. Schlag, I., Arandjelovi´c, O.: Ancient Roman coin recognition in the wild using deep learn- ing based recognition of artistically depicted face profiles. In Proc. IEEE International Conference on Computer Vision pp. 2898–2906 (2017)

16. Yue, X., Dimitriou, N., Arandjelovi´c, O.: Colorectal cancer outcome prediction from H&E whole slide images using machine learning and automatically inferred phenotype profiles. In Proc. International Conference on Bioinformatics and Computational Biology (2019)

17. Zaharieva, M., Kampel, M., Zambanini, S.: Image based recognition of ancient coins. In Proc. International Conference on Computer Analysis of Images and Patterns pp. 547–554 (2007)