Digital Applications in Archaeology and Cultural Heritage 37 (2025) e00403

Contents lists available at ScienceDirect

Digital Applications in Archaeology and Cultural Heritage

journal homepage: www.elsevier.com/locate/daach

Character recognition in Byzantine seals with deep neural networks

Théophile Rageau [a,b,*], Laurence Likforman-Sulem [b,*], Attilio Fiandrotti [c], Victoria Eyharabide [d], Béatrice Caseau [e], Jean-Claude Cheynet [f]

[a] École normale supérieure Paris-Saclay, 4 Avenue des Sciences, Gif-sur-Yvette, 91190, France
[b] Institut Polytechnique de Paris, Telecom Paris, 19 Place Marguerite Perey, Palaiseau, 91120, France
[c] Universitá di Torino, UniTo, 12 via Pessinetto, Torino, 10153, Italy
[d] Sorbonne University, STIH Laboratory, 28 rue Serpente, Paris, 75006, France
[e] Sorbonne University, 1 rue Victor Cousin, Paris, 75005, France
[f] Collège de France, CNRS, 52 rue du Cardinal Lemoine, Paris, 75231, France

ARTICLE INFO

Keywords:
Byzantine cultural heritage
Computer-based sigillography
Seal images
Character localization
Character recognition
Ancient Greek characters
Deep nets
Object detector

ABSTRACT

Seals are small coin-shaped artifacts, mostly made of lead, held with strings to seal letters. This work presents the first attempt towards automatic reading of inscribed text on Byzantine seal images. Byzantine seals are generally decorated with iconography on the obverse side and Greek text on the reverse side. Text may include the sender's name, position in the Byzantine aristocracy, and elements of prayers. Both text and iconography are precious literary sources that wait to be exploited electronically, so the development of computerized systems for interpreting seals images is of paramount importance. This work's contribution is hence a deep, two-stages, character reading pipeline for transcribing Byzantine seal images. A first deep convolutional neural network (CNN) detects characters in the seal (character localization). A second convolutional network reads the localized characters (character classification). Finally, a diplomatic transcription of the seal is provided by post-processing the two network outputs. We provide an experimental evaluation of each CNN in isolation and both CNNs in combination. All performances are evaluated by cross-validation. Character localization achieves a mean average precision (mAP) greater than 0.9 at the intersection of union threshold of 0.5. Classification of characters achieves an accuracy greater than 0.92. Such performance compares favorably to similar tasks such as the recognition of inscribed characters on ancient coins. At transcription level, we provide novel performance results in terms of Character Error Rate. This is novel for seal images and differs from results on isolated character recognition.

1. Introduction

Byzantine sigillography is the study of seals that were produced during the Byzantine period. Sigillography, as epigraphy and numismatics, is a discipline working on *inscribed* objects. The writing is different from the one used in manuscripts. Seals are small circular artifacts (made of lead or sometimes gold) that were held with a string to 'seal' official or private documents or objects. Mentions of family names, function in the State and dignities provide important information on the aristocracy and the administration of the Byzantine State. In this article we focus on Byzantine seals from the VIII–XIIth centuries. They often contain the image of a saint or a cross on the obverse side and provide information on the preferred religious motif of the sigillant (Campagnolo-Pothitou and Cheynet, 2016; Cheynet, 2019). More than 60,000 seals have been unearthed and are preserved in public or private collections worldwide (Sopracasa et al., 2020).

Seals have two sides: obverse and reverse. On the obverse side, one often finds iconography. On the reverse side, a text is usually written in Byzantine Greek capital letters, including three to seven 'text lines' depending on the size of the seal and of the letters. This text may contain the name of the sender, his/her social position, and elements of prayers. Sometimes, either side or both sides may contain a monogram that the document's reader should decipher by combining letters. Text on seals is composed of ancient Greek letters, whose glyphs have evolved over time. These variations are encoded by the Athena Ruby font (Kalvesmaki, 2015). This font was conceived specially for Byzantine seals and includes about 700 distinct glyphs.

Fig. 1 shows the obverse and reverse sides of a sample seal. The characters are made by an engraver who creates reliefs on a *boulloterion*, the tool used to strike lead, silver and gold bullae. Digitization has been performed by a camera. Notice that background and characters

* Corresponding authors.
E-mail addresses: theophile.rageau@gmail.com (T. Rageau), likforman@telecom-paris.fr (L. Likforman-Sulem), attilio.fiandrotti@unito.it (A. Fiandrotti), victoriaeyharabide@gmail.com (V. Eyharabide), beatrice.caseau@sorbonne-universite.fr (B. Caseau), jean-claude.cheynet@wanadoo.fr (J.-C. Cheynet).

https://doi.org/10.1016/j.daach.2025.e00403
Received 24 February 2023; Received in revised form 28 October 2024; Accepted 17 February 2025
Available online 5 March 2025
2212-0548/© 2025 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Fig. 1. Sample seal images. The obverse side (left) includes iconography. The reverse side (right) includes text of same color as background. Ground truth character bounding boxes are superimposed in red color. An abbreviation marker (right) is superimposed in blue color. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

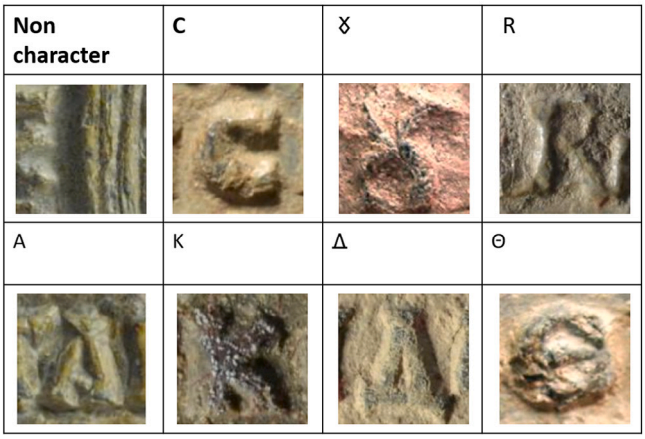

Fig. 2. Sample character images cropped from seal images. Such images are challenging for character recognition.

have the same color, and in addition shades may be present depending on the light source position during the digitization process. Therefore, traditional OCR (Optical Character Recognition) approaches cannot be applied to such images since they require contrast between foreground (characters) and background (see Fig. 2)

The scene depicted in this seal (left) is St. Nicholas in his customary episcopal garb, holding the Book of the Gospels in his left hand and blessing with his right.

Since the surface of a seal is small (between 10 to 50 mm in diameter), engravers have gained room by removing word spaces, omitting or fusing characters, and omitting even entire words. Consequently, the text is often abbreviated. As an example: the inscription on the reverse side of the seal presented in Fig. 1 +Σοφήα ϑυγάτ Μχ μαγίστρου+ actually corresponds to a longer text that starts with the "croisette" symbol (unseen characters are boxed): +Σοφήαϑυγάτηρ Μιχαὴλ μαγίστρου+, Sofía thygátir Michaíl magístrou (in Greek), which means: "Sophie, daughter of Michel magistros" (English translation).

Fig. 3 illustrates this abbreviation process in Byzantine seals. The first one corresponds to the ligature "*ου*" (two characters fused into one symbol). The pattern "*ου*" corresponds to a widespread grammatical form in Greek. The second one corresponds to one word (Michael), reduced to its first (M) and third (χ) letters.

To help historians interpret seals, the *BHAI* project proposes an innovative approach for Byzantine sigillography based on artificial

Fig. 3. Examples of text abbreviation. On the left ligature $ου$, on the right the upper mark symbolizes an abbreviated form: $M\chi$ corresponds to word $Μιχαηλ$ (Mihaïl or Michael).

intelligence.[1] The objectives of this project include seal dating, iconography interpretation (what the scene represents), and text recovery (despite its abbreviated form and damaged characters). These objectives rely on recognition tasks such as figures and objects recognition (obverse) and character recognition (reverse). This paper proposes a learnable two-stage character reading approach capable of producing a diplomatic transcription of seal images (Fig. 4). Our approach is implemented as a pipeline that includes the following stages:

- a convolutional object detector trained to detect characters from seal images (character localization). From the predicted bounding boxes, isolated character images can be cropped.
- a convolutional image classifier trained to recognize letters in the crops above (character classification)
- finding the text lines and transcribing reverse seal images from characters' positions and labels.

To our knowledge, the proposed system is the first recognition system dedicated to the recognition of seal images, in particular Byzantine seals. Our approach has been developed in the context of low resource data, since annotated seal images are scarce.

The rest of the paper is organized as follows. We present related works dealing with cultural heritage and written artifacts in Section 2. In Section 3, we describe our annotated dataset of Byzantine seals from two different collections. Our proposed two-stage approach is detailed in Section 4, including model architectures (Sections 4.1 and 4.2, respectively) and training procedures. Finally, in Section 5 we describe our experiments at reading characters from our dataset.

2. Related works

Ancient written documentary sources are important for understanding the history of a period of time in a specific geographical area. Such sources were written for the most ancient ones on argile and wood tablets (Rusakov et al., 2021; Molton et al., 2003), papyri (Pirrone et al., 2021), parchments (Dhali et al., 2020; Leydier et al., 2014), funerary inscriptions and funerary stele (Bobou et al., 2020; Andreu and Serrano, 2019), metal especially coins (Kiourt and Evangelidis, 2021), and stones (Sfikas et al., 2020). Conferences such as DATeCH and HIP (Antonacopoulos, 2019; Neudecker, 2023) highlight the interest of recognizing the text content of ancient documents.

These last decades, several projects have been launched in order to build computerized systems for automatically deciphering digitized written artifacts (ILAC for coins, Arandjelović (2010), ARCADIA for pottery sherds (Chetouani et al., 2020). There are projects and challenges focused on ancient Greek such as the Ancient Lives project (Swindall et al., 2021) and the ICDAR 2023 competition for the detection and recognition of characters on Greek papyri (Seuret et al., 2024, 2023). The best approaches of this competition were (Vu

[1] This project is lead by V. Eyharabide, in collaboration with historians, experts in Byzantine history and sigillography

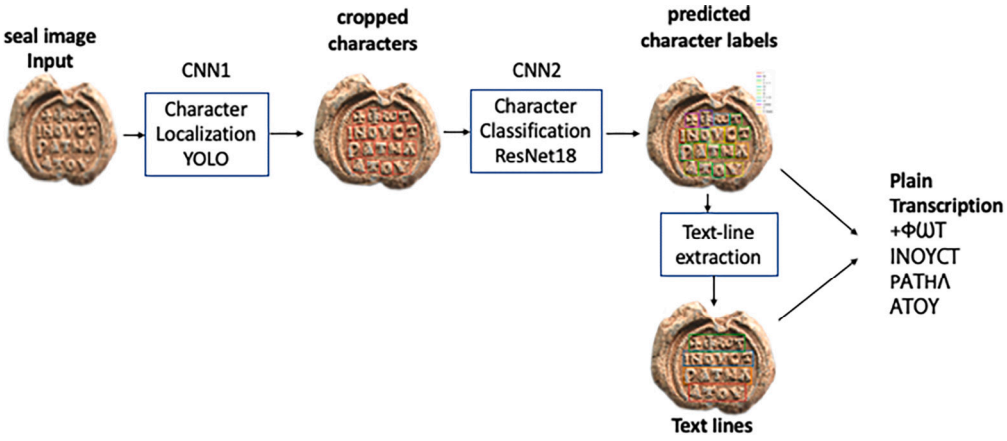

Fig. 4. Pipeline of the proposed two-stage character recognition approach. A YOLO network detects characters within each seal providing bounding boxes (object localization). Next, a ResNet18 network classifies each detected character (object classification). The output is the diplomatic transcription of the input reverse seal image.

and Beurton-Aimar, 2023; Turnbull and Mannix, 2024). These works deal with Greek handwritten characters extracted from papyri images while we deal with Greek inscribed characters. Handwritten characters in papyri are numerous and tend to touch and overlap each other, while inscribed characters lack background/foreground contrast. In Assael et al. (2022) the Ithaca system deals with ancient Greek texts collected from stone inscriptions. It predicts date and geographical location of an input text. It also predicts missing characters, assuming the number of missing characters is known. However this system does not perform character recognition.

Seal collections may be hard to interpret even for human beings, mainly due to their damaged state and the lack of ancient language knowledge and historical context (Souibgui et al., 2022). The above systems rely on machine learning approaches. Such approaches require a training step that uses annotated data. In the best cases, domain experts can annotate digitized written material to provide transcriptions, meta-data, and bounding boxes of objects of interest. Nevertheless, annotation is a time-consuming task requiring expertise and knowledge. Therefore, only a limited number of images may be annotated.

Depending on the manufacturing process, it will be more or less easy for a computer to read written artifacts. In particular, well-contrasted writings with separated characters are the easiest to recognize. Moreover, writings on coins, tablets, and seals result from a 3-dimensional manufacturing process. Digitization into 2-dimensional images with a camera provides shadows (depending on the light source and contrast) that are difficult to process. An additional difficulty is the size of the letters and the style of writing.

Scene Text recognition approaches aim to read the text in natural scenes images digitized by a camera (Zhu et al., 2016). Such approaches may also be suitable for other written artifacts. In this context, there are two main approaches: one based on machine learning and the other on deep learning (a mix of both is also possible when several tasks are required). The machine learning approach consists in extracting features (f.i. SIFT features) from keypoints, then defining a matching or classifying based approach (f.i. SVMs). Such approach is described in (Arandjelović, 2010) for classifying coins into 65 classes (portrayed Emperors). But there is no attempt to read characters. In contrast (Kavelar et al., 2014), aims at reading Emperor names from recognized characters. Depending on the lexicon and other parameters, the Word Recognition Rate ranges from 29 to 67%. The objective in (Pan and Tougne, 2017) is to read coin release dates from recognized digits. Average recognition accuracy is about 80% for digit images collected under well controlled lighting conditions.

In contrast to Machine Learning approaches, deep learning approaches automatically learn features and store them in their convolutional layers. If enough annotated data are available for training, it is possible to train multi-class systems that both localize and recognize textual elements (Bartz et al., 2017). However, data scarcity is a common problem for ancient artifacts. One solution is the generation of synthetic data (Madi et al., 2022) and fine-tuning on real data (Nguyen et al., 2021). As a result, millions of synthetic characters may be necessary. Another approach consists of first locating text regions (see Sfikas et al. (2020) for the detection of Greek inscriptions on monuments and then applying a character recognition approach in each text region. These regions can enclose a paragraph, a text line, or a character.

Fig. 5. Example of a seal image before and after preprocessing.

This two-step approach is suitable for facing data scarcity: the combined localization/recognition task is decomposed into two more straightforward tasks. As a result, two networks are created, each being lighter than a single large network.

In the present work, we face data scarcity, low-contrast images, and damaged characters. Thus, we decompose the recognition task into two stages: (i) character localization and (ii) character classification (see Sections 4.1 and 4.2).

3. Seal image dataset

Our dataset includes seal images from two collections of Byzantine seals: the Geneva Zacos collection and the Tatish collection. George Zacos collected the seals of the first collection, and her widow Janet Zacos gave a number of those seals to the "Musée d'art et d'histoire de Genève" in Switzerland. The second collection belongs to Nuri Yavuz Tatiş, and it is one of the richest and finest private seals collections in Turkey.

As shown in Fig. 5, all images were manually cropped to center the image on the seal and eliminate all irrelevant surrounding information, such as handwritten notes. The background in each image was set to white to eliminate any side default since the seals had been photographed originally on different supports. In some cases, the images

Fig. 6. Annotated seal at pixel level (reverse side). Character color codes: red = character, blue = ligature, purple = abbreviation mark. Characters are in capital letters and there is no space between words. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Character	Glyph	#Samples	Character	Glyph	#Samples
Alpha	A	182	Rho	P	133
Chi	X	23	Beta (open)	R	55
Sigma	C	141	Upsilon	V, Y	60
Gamma	Γ	54	Lambda	Λ	65
Kappa	K	88	Delta	Δ	35
Theta	Θ	54	Pi	Π	65
Eta	H	78	Epsilon	Є	113
croisette	+	82	Tau	T	123
Nu	N, И	93	Omicron	O	146
Omega	Ѡ	104	Mu	M	40
Ligature OU	Ꙋ, Ѹ	48	Ligature CT	Ҁ	9
Phi	Ф	52	Zeta	Z	8
Iota	I	154	Xi	Ѯ	4
S= KAI	S	24	Psi	Ψ	2
			Beta (closed)	B	7

Fig. 7. Ancient Greek characters with glyphs and their occurrences. The last 5 characters have fewer than 10 samples and were not considered in our experiments.

were also rotated to correctly display characters and iconographic scenes. Their size is 1000 × 1000 pixels (high resolution).

Thanks to our trained experts in Greek and Byzantine history, characters and objects on seals of the Geneva Zacos collection were manually segmented at pixel level (semantic segmentation). Our experts also provided a transcription in a standard Greek font and in the specialized Athena Ruby font (see Fig. 6). The whole annotation process took them three months.

Fig. 7 shows the 29 character classes we use in this work. Classes ξ, ψ, ζ and closed β are quite unfrequent. The Upsilon and Nu classes include two different glyphs each. Character S is the abbreviated form of word Και. There are also two ligatures: CT for στ (unfrequent) and OU for ου. When transcribed in standard Greek, the number of character classes drops to 24 classes, the number of characters in the Greek alphabet.

From the pixel-based annotations, a bounding box around each character was easily derived. The number of seals we kept in the Geneva Zacos collection is 57 (29 reverse, 28 obverse). The Tatish collection was not annotated at pixel level. Only transcriptions in a standard Greek font were provided. Thus we manually annotated 37 additional reverse and 8 obverse seal images by drawing bounding boxes around characters and assigning them a label. Such annotation is less complete compared to the Geneva Zacos collection however. A total of 102 annotated images from both collections are available to train and evaluate the proposed approach (Table 1).

Table 1
Number of annotated seal images. The annotation includes at least bounding boxes coordinates and character classes.

Collection	#reverse	#obverse	#total seal images
Zacos	29	28	57
Tatish	37	8	45
Total	66	36	102

4. Proposed approach

In this section, we describe our learnable approach in order to transcribe text present on seal images. Speaking of text detection in general, two approaches are possible. The first and most desirable approach consists in jointly detecting and classifying the characters. This approach is known as *object detection in the wild* and has been solved before for several different tasks. For example, YOLO is a family of one-shot joint object detection and classification deep convolutional architectures trainable end-to-end in a single shot. However, even if properly pretrained, such approaches usually require large annotated training sets in the order of tens of thousands of images, whereas we have far fewer annotated images in our case. Therefore, we opted for a different approach consisting of (i) localizing character bounding boxes in the image and (ii) reading out the characters previously localized as a simple image classification approach. The advantage of this second approach is that the detector needs to solve a simpler task (just localization) instead of two tasks simultaneously (localization and recognition). This approach splits a larger problem into two inherently simpler sub-problems, where each can be solved by learnable models trained over far fewer annotated samples. We conducted experiments with Yolov5. It can be noted that Yolov5 and Yolov8 were also used by the leading teams in 2023 at the "ICDAR 2023 Competition on the detection and recognition of Greek Papyri". The best approaches at this competition [Vu and Beurton-Aimar 2023] [Turnbull and Mannix 2024], also consisted in separate detection and recognition stages. Since characters are hard to recognize, dedicated classifiers are necessary.

Finally, diplomatic transcriptions are produced combining the outputs of both networks and apply a Hough-based approach that sorts the detected and classified characters into text lines.

4.1. Character localization

As a first step, we rely on the *small* version of YOLO v5 (Jocher, 2020) (You Only Look Once) to detect characters in the image (object localization). The output of YOLO is a list of tuples for each object detected in the image. Each tuple has the form (x, y, width, height), where (x, y) stands for the re-normalized coordinates of the center of the predicted bounding box and (width, height) corresponds to the re-normalized size of the bounding box. Namely, our YOLO learnable parameters have been previously initialized by training the network over the COCO (Common Objects in Context) dataset (328,000 images). We refine such a network by fine-tuning the parameters over seal images for 300 epochs. Due to the limited number of data, we massively resort to geometric transformation (image shifts, scale variations) to augment the training set diversity. We optimize network parameters (*i.e.* network weights) with SGD (Stochastic Gradient Descent) for 300 epochs (the apparently large epoch count is due to the limited size of our actual training set) and a linear learning rate scheduler from 0.01 to 0.00001.

Table 2
Number of seal images for the character localization task (top) and number of characters for the character classification task (bottom). Samples will be divided into training/test data according to a cross-validation framework.

	reverse	obverse	non chars	total
Character Localization (YOLOs v5)	66	36	–	102
Character Classification (ResNet18)	1857	456	150	2463

Table 3
Character Localization evaluation on reverse seals. Recall, precision and mAP metrics (in %) obtained by cross validating over K = 10 folds. The value of the IOU threshold equals 0.5.

Precision	Recall	mAP	
		@0.5	@[0.5:0.95]
93.74	89.80	94.40	65.78

We experimented with training batches of 2, 4, 8, and 16 images, and the best performance-complexity tradeoff was obtained with a size equal to 4. At the end of each epoch, the model is evaluated on a validation dataset (15% of the training data) and the mAP (mean average precision) and F1-score are computed (Padilla et al., 2020). The best model is then evaluated against testing data, *i.e.*, a set of images the network has not seen at training time.

4.2. Character classification

For character classification, we rely on a ResNet18 architecture (He et al., 2015) pre-trained on the ImageNet dataset (Russakovsky et al., 2014), recasting the character reading problem as a crop classification task. ResNet is still a popular model for classification due to its simple design and its efficiency. We conducted a number of preliminary experiments concluding that this architecture offers good enough classification accuracy even when trained over a limited number of annotated samples. The network is trained over isolated character crops extracted from the training images resized to 256 × 256 pixels. To cope with data scarcity, we train the classifier on reverse and obverse character images. In addition, we rely on augmentation strategies to increment the apparent training set size. The number of character classes in our setup is 24, corresponding to the most frequent classes in Fig. 7. The Croisette class was included (even if it is not a letter) to avoid confusion since its shape is similar to a character.

While the ResNet18 is trained over character-exact crops from the annotated training images, it is expected to operate on crops extracted by the YOLO-based localization pipeline when deployed in the complete localize-and-classify pipeline. These crops are expected to be either inaccurate or incorrect altogether.

Towards this end, we added an extra non-character class corresponding to bounding boxes containing no character or damaged characters. The non-character images simulates false positives detected by the previous localization stage.

For all characters, we provide their ground truth bounding box coordinates, except for the non-character class, which has a bounding box size equal to the mean character size of the seal image to which it belongs. In summary, 150 non-character images are randomly generated at each iteration of the K fold cross-validation processes and added to the character image set.

Table 2 shows the number of seals used for the localization task. Good quality obverse images are added for augmenting the data. For the classification task, we provide the number of characters used for this task. It can be noted that these characters are cropped from the seal images used for localization. Folds at character level (classification task) follow the folds at seal level (localization task). This ensures that training and test characters always belong to distinct seals (see Section 5.1).

4.3. Text transcription

Text lines are extracted from bounding boxes and label predictions using a Hough-based approach described in Likforman-Sulem et al. (1995). This approach consists in detecting the most likely lines of text (the ones that intersect the most bounding boxes) following a hypothesis-validation strategy which is iteratively activated until the end of the segmentation is reached. At each stage of the process, the best text-line hypothesis is generated in the Hough domain. Afterwards, the line's validity is checked in the image domain using proximity criteria. Finally, a diplomatic transcription is obtained by combining the previously detected lines with character labels (lines of the transcription correspond to those of the input image).

5. Experiments

We describe in the following, the evaluation protocols for each task. We then evaluate each task in isolation, and in combination for the entire pipeline.

5.1. Evaluation protocol

We adopt a cross validation framework with K folds, *i.e.* we randomly draw K folds out of the annotated data and in turn we leave out one fold for testing while we train on the remaining K-1 folds. All the results presented below are relative to the reverse side of seals and they all represent the mean of the K folds performance. For the localization task, the folds include seal images and for the classification task, characters. To avoid including characters from the same seal in training and testing folds, we adopt the following approach. First, seal images are randomly placed into ten folds (K = 10 folds). Since there are 102 seal images, there is about ten images per fold. Second, in a given fold, characters are cropped from the seal images placed in this fold. This ensures that test character samples and training samples do not belong to the same seals. Fig. 9 shows the character frequencies for two configurations of training folds in the Cross-Validation process. Since the characters on the reverse images include rare letters, and our dataset is scarce, we have augmented our data by adding to the training sets of both tasks, a few observe seal images and characters.

In Section 5.3, the classification task is evaluated from ground truth bounding boxes. However when we will evaluate the complete pipeline (see Section 5.4), we will consider the bounding boxes provided by the first CNN (YOLO, character localizer).

5.2. Character localization

To validate the YOLO system used as a character localizer, we use cross-validation with $K = 10$ folds. Fig. 8 shows the evolution of the training and validation losses (two columns on the left) and the metrics on a validation set (two columns on the right) over epochs. The validation set includes 15% of the training data.

Fig. 8 shows that the mAP@[0.5:0.95] reaches 0.7 after 300 epochs which shows that our model is confident when predicting bounding boxes. In fact, a high value of this metric indicates that even for high values of the IoU (Intersection Over Union) threshold (for instance $IoU_t = 0.9$ or 0.95), the model achieves decent results. For obtaining the cross-validation performance, we cycle over the folds. For each iteration of the cross validation, the best model on the validation set is recorded and tested on the test fold. Table 3 provides Cross validation performance on reverse images for three popular metrics (recall, precision, mean average precision mAP).

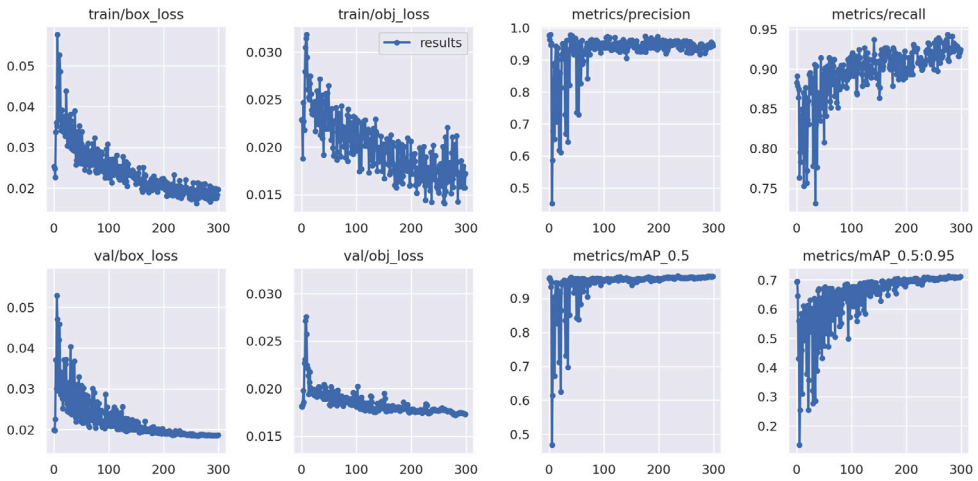

Fig. 8. Training (top) and validation (bottom) loss curves (two columns on the left), and evolution of recall, precision and mAP metrics according to the number of epochs (two columns on the right).

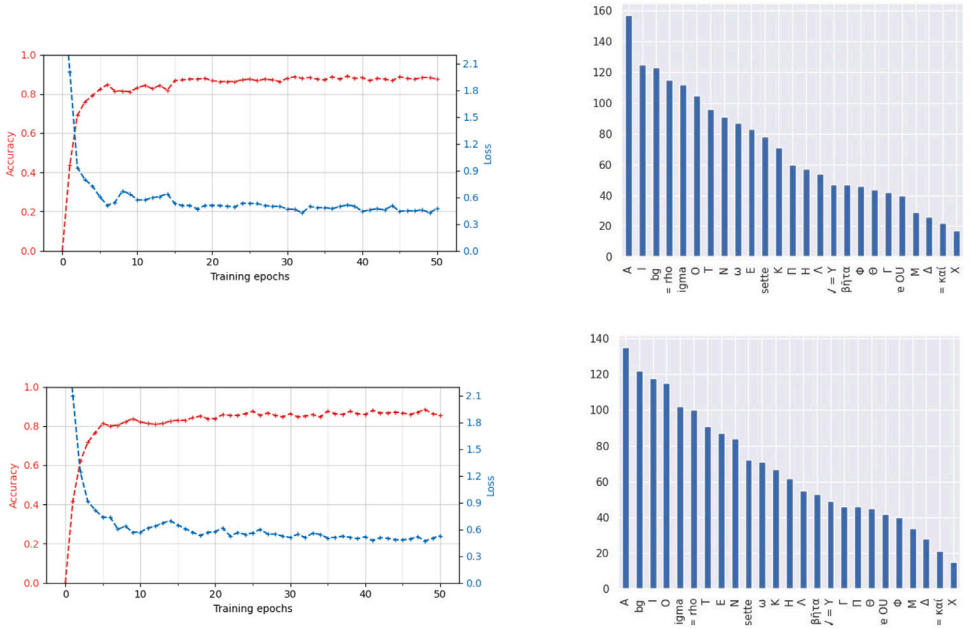

Fig. 9. (left) Validation Accuracy and Loss curves, for the ResNet classification task and two Cross-Validation steps. (right) Character-class frequencies in the corresponding training folds.

5.3. Character classification

Next, we evaluate the performance of the ResNet18 network in charge of classifying the characters localized by the YOLO character detector. We decided to train our model on ground truth crops rather than on the deep network detector predictions because they might be inaccurate compared to the ground truths and might introduce a bias in our dataset. We also add padding around each crop in the training set to make the network more robust for predicting on new data. The results in Table 4 are relative to the ResNet18 network evaluated on ground truth crops for the 20 most represented classes (which are composed at least of 50 samples) plus the non-character class defined in Section 4.2.

Confusion matrices do not include rare character classes as this would result in folds that lack even a single class-representative sample. We need approximately 5 character samples per fold for each class. Therefore, a total of about 50 characters per class is necessary. Table 4 provides cross-validation accuracies and F1-scores for the classification task computed on the reverse side of test images. We trained the classification model from pretrained weights learned on the ImageNet dataset (Russakovsky et al., 2014). We used some common data augmentation such as affine transformation and change of color. For each iteration of the cross-validation procedure, we train a model for 100 epochs with early stopping. For each training, the best model in terms of Top-1 accuracy and F1-score on the validation set (15% of the training set) is recorded and tested on the test fold. Fig. 9 plots the loss curve and the validation accuracy curve for two training iterations of the cross-validation procedure.

Fig. 10 shows the confusion matrices (Top-1 on the left and Top-3 on the right) obtained after one single training of the cross-validation

Table 4
Character classification accuracies and F1-Score obtained by cross-validation for $K = 10$ folds. All metrics are provided in percentages (%).

Top-1 acc.	Top-2 acc.	Top-3 acc.	F1 Score
92.12	96.10	97.37	91.65

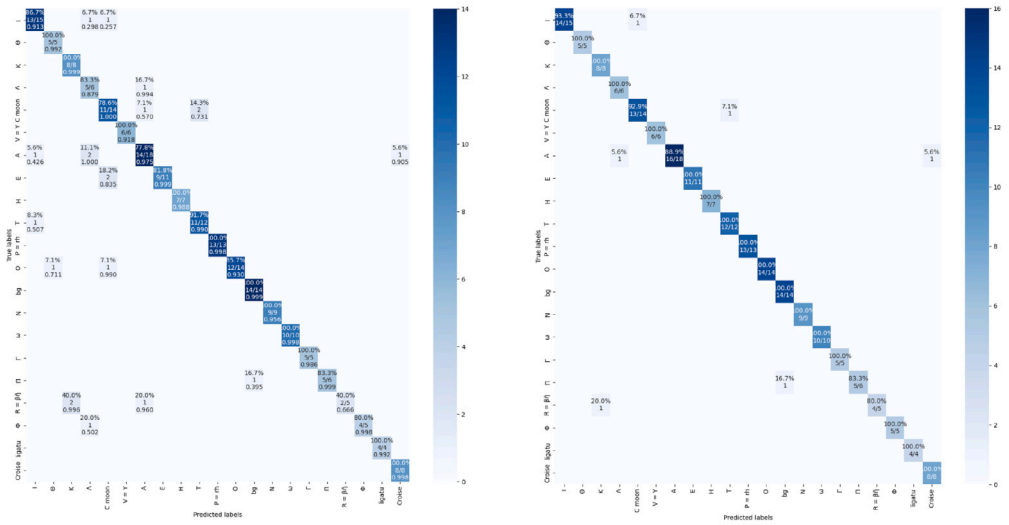

Fig. 10. Top-1 (left) and Top-3 (right) confusion matrices for the character classification task, on one test fold. We consider the 20 most represented (at least 50 samples) character classes, plus one pseudo-class for false positives.

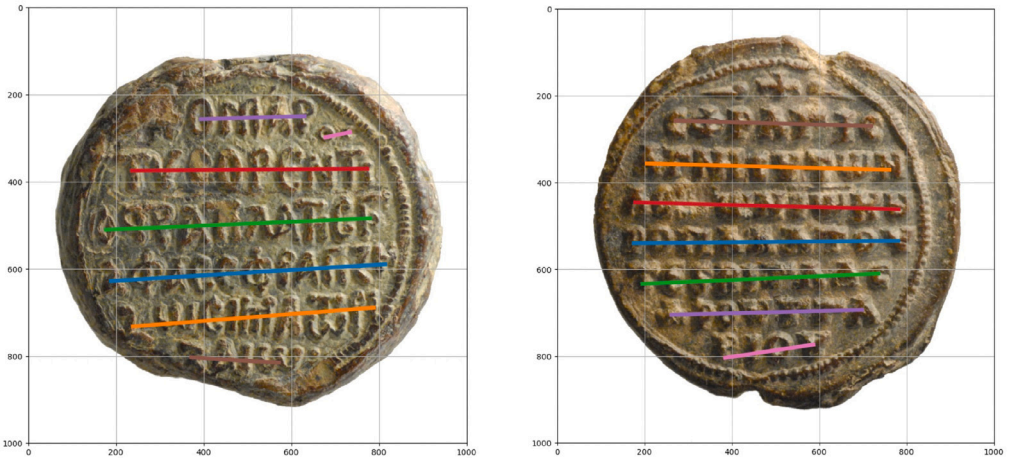

Fig. 11. Text lines extracted from sample seal images.

procedure. The Top-1 confusion matrix (left) shows that the model achieves decent results but still makes some mistakes. The Top-3 confusion matrix is almost diagonal, it indicates that the model has understood well how to dissociate the classes and makes only some mistakes on the hard examples.

We have measured the inference time of our PyTorch implementation on a typical GPU (NVIDIA V100 16 GB). The detection of characters with YOLO takes about 0.5s per seal image. The classification task with ResNet takes about 40 millisecond per character. Thus processing a whole seal image takes about 0.9 s, assuming that ten characters are inscribed.

5.4. Transcription evaluation

In the previous section, the pipeline classification performance has been evaluated from ground truth bounding boxes. However, bounding boxes predicted by the YOLO character detector may be inaccurate and thus differ from the ground truth. Several cases explain such difference:

- the bounding boxes (ground truth/predicted) overlap but are not identical (their IOU is greater than a threshold (f.i. 0.5))
- a bounding box is predicted at a location where there is no character (False positive)
- a character has been missed (no bounding box predicted, or IOU less than 0.5)

In the first case, even if bounding boxes are not identical, the ResNet18 character classifier could recognize the character cropped with the predicted box. Thus we evaluate the pipeline through the predicted transcription, the sequence of characters output, whatever their precise location. However, the output of the application of ResNet18 to all the bounding boxes detected by YOLO is a set of class score distributions. To obtain a diplomatic transcription, these characters are first grouped into text-lines (see Section 4.3). Then, they are ordered within each line.

Fig. 11 shows sample seals and their extracted text lines.

To compare the predicted character sequence of one seal with its ground truth sequence, we adopt the CER metric (Character Error Rate). This metric relies on the Levenshtein distance. It is equal to the minimum number of characters that must be deleted, inserted or substituted to move from the predicted to the ground truth sequence. To obtain the CER, we divide this distance by the number of characters in the ground truth sequence.

$$CER = \frac{S + D + I}{N}$$

where S is the number of character substitutions, I the number of insertions, and D the number of deletions. N is the number of characters in

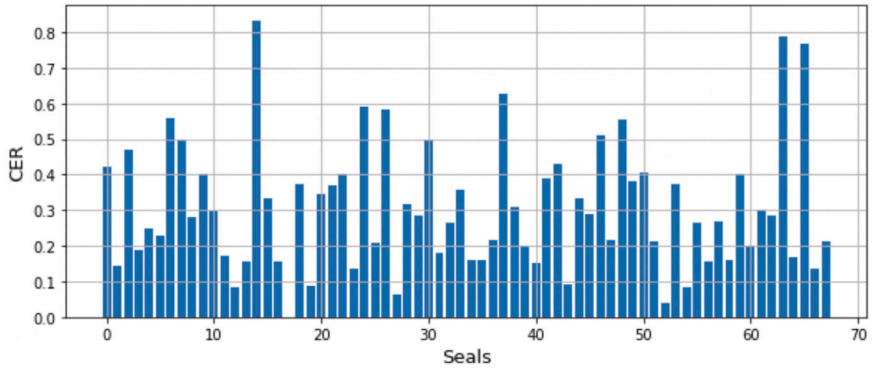

Fig. 12. Character Error Rates obtained for each seal reverse image for the complete pipeline.

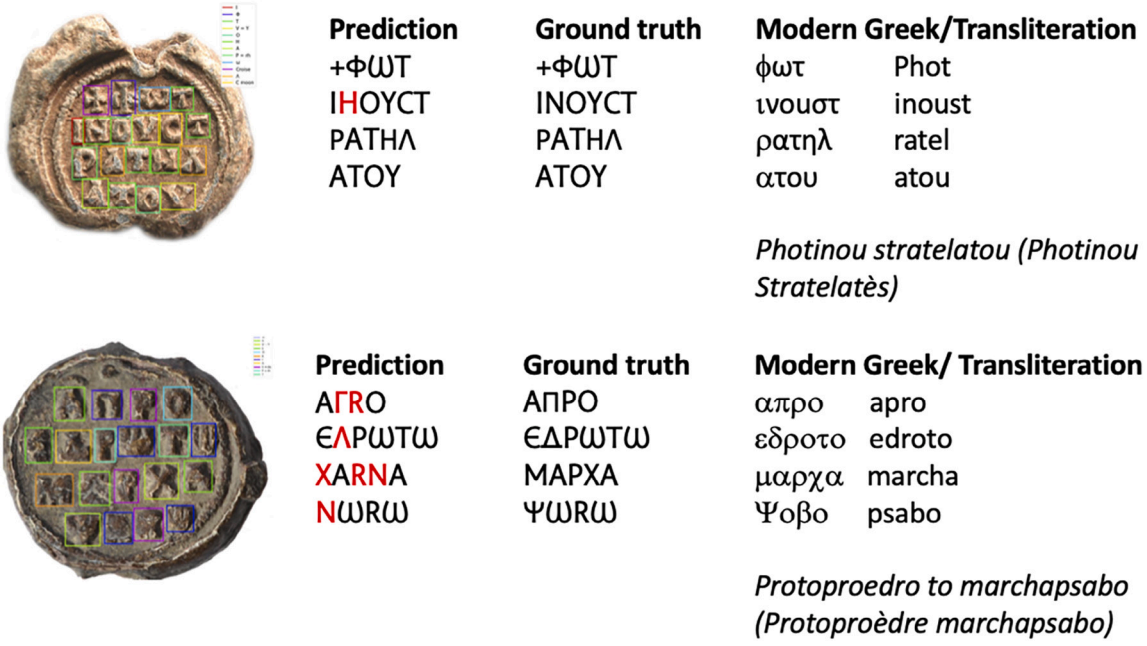

Fig. 13. Transcriptions obtained for two samples seals. Characters in red denote an error. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 5
Cross-validation results (K=10) for the complete pipeline (transcription task). We provide the CER (Character Error Rate) for each fold, and the overall performance.

Fold	1	2	3	4	5	6	7	8	9	10	overall
CER	0.32	0.27	0.30	0.33	0.29	0.29	0.34	0.25	0.24	0.39	**0.31**

the seal ground truth sequence. Following the K-fold cross-validation framework, we compute for each seal of the testing fold, its CER, and average over the seals to obtain the CER associated to this fold. The cross-validation CER is obtained by averaging over the folds (see Table 5). The Cross-validation CER equals 0.31, which proves the efficiency of the proposed approach for seals' written artifacts. In Fig. 12, we show all seal image CER values. High values of CER correspond to hard seal examples, either damaged or containing illegible characters. In this case, the detector results are less accurate (bounding boxes that do not cover all characters or the presence of false positives). The smallest CER values correspond mostly to classification errors (substitution between two closely shaped characters).

Fig. 13 shows the transcriptions obtained for two sample seals. In both seals, characters have been correctly detected by the deep localizer network. Characters have been almost perfectly recognized for the top seal. There are a few errors in the second one. Errors are due to confusions between characters. It can be noticed that characters of the second seal are much less legible, compared to the first seal. In this figure, we have added the ground truth in modern Greek, its transliteration, and the underlying text provided by an expert. The first *α* in the second seal is the abbreviation of word 'proto'. A *Stratelatès* is an army general. A *Protoproèdre* is a byzantine dignitary (X-XIth centuries), and Marchapsabo, the name of a well-known family.

Fig. 14 shows typical difficult cases. Decorative elements (o-shaped, leaves) can be seen as characters. The large letter 'M' is split into its compound elements.

6. Discussion

An overall character error rate of about 30% cannot be directly compared with high accuracies reported on standard datasets (such as the digit MNIST dataset (Lecun et al., 1998)) with deep networks. Reported errors consist of confusion between classes only. Moreover, such sets include thousands of samples, so there is no scarce data problem. Characters are also well-contrasted and correctly cropped; thus, there are no occluded or missing strokes. In contrast, the character error rate in our overall pipeline reports several errors from the

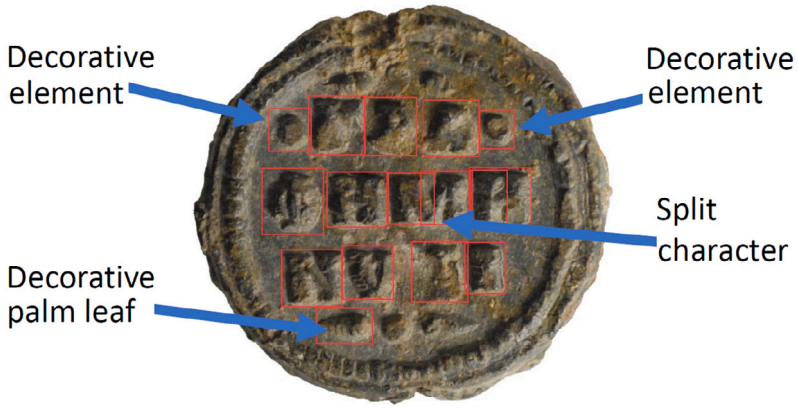

Fig. 14. Hard case samples. Localization errors due to decorative elements. Large characters such as 'M' may be split.

predicted character sequence. Errors include confusions but also false positive and false negative crops. Confusion errors may be due to non-perfect crops with missing strokes, which is not the case with standard character recognition datasets.

A fairer comparison may be made in the numismatics domain where characters/digits on coins are also inscribed on small round-shaped objects. The CoinNUMs_pcgs_m and CoinNUMs_pcgs_a datasets collected by Pan and Tougne (2017) include digit images manually or automatically cropped from coin images, respectively. The foreground and background share the same hue (as in our case), and there may be crop errors in the automatic setting. Performance for manually cropped characters is greater than for automatically cropped ones (43.2% accuracy versus 29.4% using the mean over classes of the per-class accuracy). This can be explained by the fact that automatic cropped characters include occluded strokes, or are not centered. For manually cropped characters, the relatively low performance can be explained by the small size of the digit dataset, and the use of a too basic neural network-based classifier. Using the same metric, our average per-class accuracy is equal to 89% considering the most frequent classes (20 classes) and the non-character class. Thus at character level, our performance compares favorably to similar tasks such as the recognition of inscribed characters on ancient coins.

Systems dealing with ancient Greek handwritten characters such as in Turnbull and Mannix (Turnbull and Mannix, 2024) and Swindall et al. Swindall et al. (2021) report a 89.1% character classification accuracy and a 92.7% accuracy respectively. This last performance is comparable to our 92.1% accuracy for the recognition of seal inscribed characters. However comparing these performances is not straightforward since datasets differ and characters may be cropped differently (using bounding box prediction or ground truth coordinates).

At transcription level, we provide novel performance results in terms of CER. This is novel for seal images and differs from results on isolated character recognition.

7. Conclusion and perspectives

In this work, we have proposed a two-stage approach for reading Byzantine seals.[2] A first deep neural network localizes the characters within the seal, whereas a second network classifies the previously localized characters. A character-level diplomatic transcription is then obtained by reordering characters into text lines. This transcription is the first step towards language processing to extract words, named entities, and recover the underlying text. Our approach can also be applied to the numismatics domain where characters are also inscribed on small circular objects. Due to the lack of foreground/background contrast, two-dimensional seal images are hard to decipher. Thus three-dimensional capture and representation of seals will improve character detection and recognition (Horache et al., 2021).

However, our architecture presents several constraints that we are planning to overcome. First, our system relies on a two-stages (localize and classify) approach since we do not have enough annotated images to train a single architecture performing joint character localization and classification. Towards this end, we have achieved encouraging preliminary results by boosting our training set with computer-generated seal images. Whether synthetic images (obtained f.i. by style transfer or generative AI) can replace real images in training is the object of our current endeavors. We will also include in our future work, recent versions of YOLO such as Yolov11 released in September 2024. Yolo models released by Ultralytics as somewhat incremental improvements over Yolov5, share the same basic principles. The benchmarks from Ultralytics github repository (https://github.com/ultralytics/ultralytics) shows improvements in standard object detection tasks such as MS-COCO. However, such numbers hold only under the assumption that a rich enough training set to teach the models is available, whereas the challenge we are tackling is the scarcity of annotated training samples. Thus, we will also collect and/or synthesize seal and character images for developing an end-to-end approach. At present, even the *small* version of YOLO cannot be successfully trained for the joint localization and classification task. Given the scarcity of data in this domain, a number of approaches dealing with the scarcity of samples have been or are being evaluated in the literature even beyond collecting more seal and character images. To avoid overfitting, lightweight dedicated models could be combined using an ensemble approach.

A ResNet classification system could not be learnt from 2,500 character samples only. Thus ResNet is pre-trained from more than 1 million ImageNet images, then fine tuned with seal character samples in the training folds. Instead of ImageNet images, Greek character images could be used for the pre-training, using the Ancient Lives datasets (Swindall et al., 2021).

Synthesizing seal images by an AI generative text-to-image approach (Reed et al., 2016) may be in the future a solution in order to build a corpus of degraded character images. Starting from a detailed and adequate prompt, it will be possible in a near future to generate realistic images with desired properties. Second, in this work, we have restrained ourselves from transcribing the reverse side of the seals where the text is oriented on horizontal lines. The obverse side may include text oriented on a curve and is often cluttered with images, hindering the transcriptions. Our future research also includes transcribing the obverse side by properly detecting the character's orientation and location within the seal. Third and last, inscriptions on seals often correspond to abbreviated texts that the expert readers of that time were supposed to be able to decipher. Furthermore, seals often present

[2] sample characters and trained neural network weights will be released at https://github.com/LIK50/Byzantine_chars.

damages that result in the loss of one or more consecutive letters. Natural language processing approaches may help reveal the original underlying text: for instance, in Assael et al. (2022), the damaged text is completed by a deep neural approach.

CRediT authorship contribution statement

Théophile Rageau: Writing – review & editing, Software, Methodology. **Laurence Likforman-Sulem:** Writing – original draft, Supervision, Methodology. **Attilio Fiandrotti:** Writing – original draft, Supervision, Software, Methodology. **Victoria Eyharabide:** Writing – original draft, Supervision, Methodology, Funding acquisition, Data curation. **Béatrice Caseau:** Writing – original draft, Validation, Supervision, Methodology, Investigation, Data curation. **Jean-Claude Cheynet:** Writing – review & editing, Validation, Supervision, Methodology, Data curation, Conceptualization.

Funding

This work has been partially funded by the French National Research Agency (ANR), grant number ANR-21- CE38-0001.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Andreu, J., Serrano, P., 2019. Contributions of the digital photogrammetry and 3D modelling of Roman inscriptions to the reading of damaged tituli: An example from the Hispania Tarraconensis (Castiliscar, Saragossa). Digit. Appl. Archaeol. Cult. Herit. 12, e00091.

Antonacopoulos, A. (Ed.), 2019. DATeCH2019: Proceedings of the 3rd International Conference on Digital Access to Textual Cultural Heritage. Association for Computing Machinery, New York, NY, USA.

Arandjelović, O., 2010. Automatic attribution of ancient Roman imperial coins. In: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. pp. 1728–1734.

Assael, Y., Sommerschield, T., Shillingford, B., 2022. Restoring and attributing ancient texts using deep neural networks. Nature 603, 280–283.

Bartz, C., Yang, H., Meinel, C., 2017. STN-OCR: A single neural network for text detection and text recognition. URL https://arxiv.org/abs/1707.08831. (Accessed on 20 April 2024).

Bobou, O., Kristensen, N.B., McAvoy, S., Raja, R., 2020. Archive archaeology in Palmyra, Syria a new 3D reconstruction of the Tomb of Hairan. Digit. Appl. Archaeol. Cult. Herit. 19, e00164.

Campagnolo-Pothitou, M., Cheynet, J.-C., 2016. Sceaux de la collection Georges Zacos au Musée d'Art et d'Histoire de Genève. Musée d'art et d'histoire de Genève.

Chetouani, A., Treuillet, S., Exbrayat, M., Jesset, S., 2020. Classification of engraved pottery sherds mixing deep-learning features by compact bilinear pooling. Pattern Recognit. Lett. 131, 1–7.

Cheynet, J.-C., 2019. Les sceaux byzantins de la collection Yavuz Tatiş. ZeroBooks.

Dhali, M.A., Jansen, C.N., de Wit, J.W., Schomaker, L., 2020. Feature-extraction methods for historical manuscript dating based on writing style development. Pattern Recognit. 131, 413–420.

He, K., Zhang, X., Ren, S., Sun, J., 2015. Deep residual learning for image recognition. http://dx.doi.org/10.48550/ARXIV.1512.03385, URL https://arxiv.org/abs/1512.03385. (Accessed on 10 May 2024).

Horache, S., Deschaud, J.-E., Goulette, F., Gruel, K., Lejars, T., Masson, O., 2021. Riedones3D: a celtic coin dataset for registration and fine-grained clustering. arXiv:2109.15033.

Jocher, G., 2020. YOLOv5. https://github.com/ultralytics/yolov5. (Online; Accessed on 25 April 2022).

Kalvesmaki, J., 2015. Introducing Athena Ruby, Dumbarton Oaks' new font for Byzantine inscriptions. In: Rhoby, A. (Ed.), Inscriptions in Byzantium and beyond: Methods – Projects – Case Studies. pp. 121–126.

Kavelar, A., Zambanini, S., Kampel, M., 2014. Reading the legends of Roman Republican coins. ACM J. Comput. Cult. Herit. 7 (1), 5:1–5:20. http://dx.doi.org/10.1145/2583115.

Kiourt, C., Evangelidis, V., 2021. AnCoins: Image-based automated identification of ancient coins through transfer learning approaches. In: Del Bimbo, A., Cucchiara, R., Sclaroff, S., Farinella, G.M., Mei, T., Bertini, M., Escalante, H.J., Vezzani, R. (Eds.), Pattern Recognition. ICPR International Workshops and Challenges. Springer International Publishing, Cham, pp. 54–67.

Lecun, Y., Bottou, L., Bengio, Y., Haffner, P., 1998. Gradient-based learning applied to document recognition. Proc. the IEEE 86, 2278–2323.

Leydier, Y., Eglin, V., Bres, S., Stutzmann, D., 2014. Learning-free text-image alignment for medieval manuscripts. In: 14th International Conference on Frontiers in Handwriting Recognition, ICFHR 2014, Crete, Greece, September 1-4, 2014. IEEE Computer Society, pp. 363–368.

Likforman-Sulem, L., Hanimyan, A., Faure, C., 1995. A hough based algorithm for extracting text lines in handwritten documents. In: Proceedings of 3rd International Conference on Document Analysis and Recognition, Vol. 2, pp. 774–777.

Madi, B., Alaasam, R., Droby, A., El-Sana, J., 2022. git-GAN: historical style transfer GAN for generating historical text images. In: Uchida, S., Smith, E.H.B., Eglin, V. (Eds.), Document Analysis Systems - 15th IAPR International Workshop, DAS 2022, la Rochelle, France, May 22-25, 2022, Proceedings. In: Lecture Notes in Computer Science, Vol. 13237, Springer, pp. 523–537.

Molton, N., Pan, X., Brady, M., Bowman, A.K., Crowther, C., Tomlin, R.S.O., 2003. Visual enhancement of incised text. Pattern Recognit. 36 (4), 1031–1043.

Neudecker, C., 2023. Proceedings of the 7th International Workshop on Historical Document Imaging and Processing, HIP@ICDAR 2023, San Jose, CA, USA, August 25-26, 2023. ACM.

Nguyen, T.-N., Burie, J.-C., Le, T.-L., Schweyer, A.-V., 2021. On the use of attention in deep learning based denoising method for ancient cham inscription images. In: Lladós, J., Lopresti, D., Uchida, S. (Eds.), 16th International Conference on Document Analysis and Recognition, ICDAR 2021, Lausanne, Switzerland, September 5-10, 2021, Proceedings, Part I. In: Lecture Notes in Computer Science, Vol. 12821, Springer, pp. 400–415.

Padilla, R., Netto, S.L., da Silva, E.A.B., 2020. A survey on performance metrics for object-detection algorithms. In: 2020 International Conference on Systems, Signals and Image Processing, IWSSIP 2020, NiterÓi, Brazil, July 1-3, 2020. IEEE, pp. 237–242.

Pan, X., Tougne, L., 2017. A New Database of Digits Extracted from Coins with Hard-to-Segment Foreground for OCR Evaluation. Front. Inf. Commun. Technol. 4.

Pirrone, A., Beurton-Aimar, M., Journet, N., 2021. Self-supervised deep metric learning for ancient papyrus fragments retrieval. Int. J. Doc. Anal. Recognit. 24 (3), 219–234.

Reed, S.E., Akata, Z., Yan, X., Logeswaran, L., Schiele, B., Lee, H., 2016. Generative adversarial text to image synthesis. In: Balcan, M., Weinberger, K.Q. (Eds.), Proceedings of the 33nd International Conference on Machine Learning, ICML 2016, New York City, NY, USA, June 19-24, 2016. In: JMLR Workshop and Conference Proceedings, Vol. 48, JMLR.org, pp. 1060–1069, URL http://proceedings.mlr.press/v48/reed16.html.

Rusakov, E., Somel, T., Müller, G.G.W., Fink, G.A., 2021. Embedded attributes for cuneiform sign spotting. In: Lladós, J., Lopresti, D., Uchida, S. (Eds.), 16th International Conference on Document Analysis and Recognition, ICDAR 2021, Lausanne, Switzerland, September 5-10, 2021, Proceedings, Part II. In: Lecture Notes in Computer Science, Vol. 12822, Springer, pp. 291–305.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A.C., Fei-Fei, L., 2014. ImageNet large scale visual recognition challenge. http://dx.doi.org/10.48550/ARXIV.1409.0575, URL https://arxiv.org/abs/1409.0575. (Accessed on 10 May 2024).

Seuret, M., Marthot-Santaniello, I., White, S.A., Serbaeva Saraogi, O., Agolli, S., Carrière, G., Rodriguez-Salas, D., Christlein, V., 2023. ICDAR 2023 competition on detection and recognition of greek letters on Papyri. In: Fink, G.A., Jain, R., Kise, K., Zanibbi, R. (Eds.), Document Analysis and Recognition - ICDAR 2023. Springer Nature Switzerland, pp. 498–507.

Seuret, M., Marthot-Santaniello, I., White, S.A., Serbaeva Saraogi, O., Carrière, G., Rodriguez-Salas, D., Christlein, V., 2024. ICDAR2023 competition on detection and recognition of Greek letters on Papyri. URL https://lme.tf.fau.de/competitions/2023-competition-on-detection-and-recognition-of-greek-letters-on-papyri/. (Accessed on 20 April 2024).

Sfikas, G., Giotis, A.P., Retsinas, G., Nikou, C., 2020. Quaternion generative adversarial networks for inscription detection in Byzantine monuments. In: Bimbo, A.D., Cucchiara, R., Sclaroff, S., Farinella, G.M., Mei, T., Bertini, M., Escalante, H.J., Vezzani, R. (Eds.), Pattern Recognition. ICPR International Workshops and Challenges, January 10-15, 2021, Proceedings, Part VII. In: Lecture Notes in Computer Science, Vol. 12667, Springer, pp. 171–184.

Sopracasa, A., Filosa, M., Stoyanova, S., 2020. The digital enhancement of a discipline. Byzantine sigillography and digital humanities. Magazén 101–128.

Souibgui, M.A., Biten, A.F., Dey, S., Fornés, A., Kessentini, Y., Gómez, L., Karatzas, D., Lladós, J., 2022. One-shot compositional data generation for low resource handwritten text recognition. In: IEEE/CVF Winter Conference on Applications of Computer Vision, WACV 2022, Waikoloa, HI, USA, January 3-8, 2022. IEEE, pp. 2563–2571.

Swindall, M., Croisdale, G., Hunter, C., Keener, B., Williams, A., Brusuelas, J., Krevans, N., Sellew, M., Fortson, L., Wallin, J., 2021. Exploring learning approaches for ancient Greek character recognition with citizen science data. In: Proceedings - IEEE 17th International Conference on EScience, EScience 2021. In: 2021 IEEE 17th International Conference on eScience (eScience), Institute of Electrical and Electronics Engineers Inc, pp. 128–137.

Turnbull, R., Mannix, E.J., 2024. Detecting and recognizing characters in Greek Papyri with YOLOv8, DeiT and SimCLR. CoRR abs/2401.12513, arXiv:2401.12513. (Accessed on 12 May 2024).

Vu, M.T., Beurton-Aimar, M., 2023. PapyTwin net: a twin network for greek letters detection on ancient papyri. In: Proceedings of the 7th International Workshop on Historical Document Imaging and Processing. HIP '23, Association for Computing Machinery, New York, NY, USA, pp. 43–48.

Zhu, Y., Yao, C., Bai, X., 2016. Scene text detection and recognition: recent advances and future trends. Front. Comput. Sci.: Sel. Publ. Chin. Univ. 10, 19–36.