EpiSignum: Taxonomy and Annotated Dataset for Symbol Recognition in Late Antique Latin Inscriptions

Alessandro Locaputo[1,*], Martina Siderini[3], Davide Redaelli[2], Andrea Brunello[2], Nicola Saccomanno[1], Stefano Magnani[2] and Giuseppe Serra[1]

[1]*Department of Mathematics, Computer Science, and Physics, University of Udine, Italy*
[2]*Department of Humanities and Cultural Heritage, University of Udine, Italy*
[3]*Technical University of Denmark, Kongens Lyngby, Denmark*

Abstract

Inscriptions are a key source for understanding the cultural and religious history of Late Antiquity, where non-textual symbols played a crucial role in early Christian visual communication. Yet, in most corpora, only transcriptions are provided; even where digitized images exist, symbol metadata is inconsistent and rarely includes precise localization. In this work, we present *EpiSignum*, a manually annotated dataset of Late Antique Latin inscriptions, accompanied by a taxonomy for the classification of recurring symbols. Developed through collaboration between computer scientists and humanities scholars, *EpiSignum* records both symbol categories and precise locations, enabling computational analyses beyond text-only transcriptions and advancing research in early Christian iconography. To establish a reference for future work, we evaluate two strong modern object detectors (YOLOv8 and YOLO-World) in zero-shot settings and fine-tune YOLOv8 on the proposed dataset, demonstrating the limitations of general-purpose models in this specialized domain. *EpiSignum* is publicly available at https://github.com/AI4CH-UniUD/EpiSignum to support reproducible research in epigraphic retrieval, indexing, and early Christian iconography.

Keywords

Latin Inscriptions, Object Detection, Computer Vision, Late-Antique Latin Inscriptions, Symbol Recognition

1. Introduction

Recent advances in machine learning, along with the large-scale digitization of previously physical-only corpora, have led to a growing interest in Digital Humanities. Epigraphy, the study of inscriptions engraved on durable materials, has benefited from projects such as the EAGLE [1] and the Epigraphik-Datenbank Clauss/Slaby [2]. These projects have supported the development of deep learning models to analyze, classify, and interpret epigraphic data, showing how such tools can help ancient historians restore and attribute inscriptions [3, 4].

Inscriptions are one of the primary sources of information for understanding ancient civilizations, their cultures, and the functioning of their society. This is particularly true for Late Antique Latin inscriptions, through which it is possible to acquire valuable insights into religious practices and beliefs between the third and the seventh centuries CE.

A distinctive feature of Christian inscriptions is the frequent presence of non-textual elements in the form of symbols, such as crosses, anchors, chrismons, doves, and others (see Figure 1). These symbols were not merely decorative; they formed a visual communication system central to expressing the religious, cultural, and social identity of the first Christian communities. In Late Antique funerary inscriptions, such symbols often had a dual function: they conveyed messages of faith and hope, and at

VIPERC2025: 4th International Conference on Visual Pattern Extraction and Recognition for Cultural Heritage Understanding, 3-4 December 2025

*Corresponding author.
locaputo.alessandro@spes.uniud.it (A. Locaputo); s252978@dtu.dk (M. Siderini); redaelli.davide@uniud.it (D. Redaelli); andrea.brunello@uniud.it (A. Brunello); nicola.saccomanno@uniud.it (N. Saccomanno); stefano.magnani@uniud.it (S. Magnani); giuseppe.serra@uniud.it (G. Serra)
0000-0003-1962-115X (A. Locaputo); 0000-0003-2063-218X (A. Brunello); 0000-0001-5916-3195 (N. Saccomanno); 0000-0001-7869-344X (S. Magnani); 0000-0002-4269-4501 (G. Serra)

© 2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings ceur-ws.org ISSN 1613-0073
published 2026-10-15

CEUR-WS.org/Vol-4295/paper7.pdf

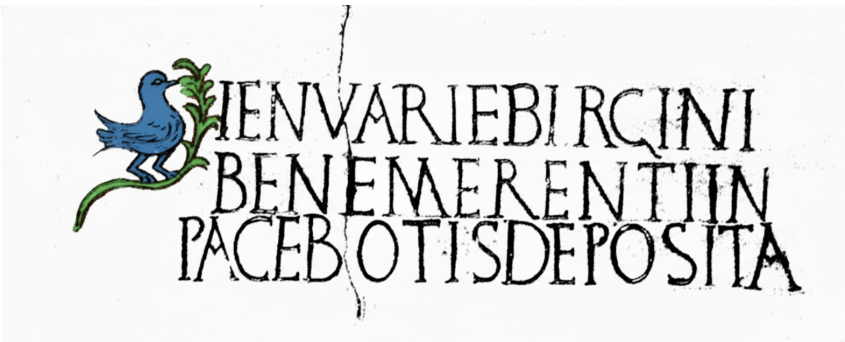

Figure 1: Example of a Late Antique Latin inscription. A dove and olive branch (*columba vel ramus olivae*). Facsimile of the original in the Galleria Lapidaria Vaticana (ICUR IX 23984).

the same time, they identified the deceased as Christian, sometimes in a discreet manner, particularly during periods when Christianity was not yet the official religion of the Roman Empire.

In most of the available epigraphic corpora, information about symbols is often omitted or included only in the textual transcription of the inscription. Even when such information is present, inconsistencies in notation or naming conventions can hinder the ability to effectively search for inscriptions containing a specific symbol [5]. More importantly, these corpora typically lack precise data regarding the placement of symbols within the inscriptions.

Creating a dataset that includes these features can facilitate the training and evaluation of machine learning models for their automatic recognition and indexing. However, this requires detailed annotation, which is a time-consuming process that relies on the involvement of domain experts in epigraphy, particularly those with experience in the iconographic analysis of Christian inscriptions.

In this paper, we contribute both a conceptual and a practical resource for the study of Late Antique Latin inscriptions. First, we introduce an original taxonomy of symbols, designed to standardize terminology and enable consistent description. Building on this taxonomy, and in collaboration with epigraphy specialists from the Department of Humanities and Cultural Heritage at the University of Udine, we manually annotated a collection of inscription images, recording both the occurrence and precise position of the relevant symbols. The taxonomy and the resulting dataset, referred to as *EpiSignum*, are presented in Section 3. Finally, with the aim of establishing a reference for future work, we provide in Section 4 the results obtained by applying YOLOv8 [6] and YOLO-World [7], two state-of-the-art models for object detection and classification, to our dataset.

2. Related work

Traditionally, collections such as the Corpus Inscriptionum Latinarum (CIL) and L'Année épigraphique have been the primary sources for Latin inscriptions [8]. In recent times, several projects have developed digital collections. For example, the Electronic Archive of Greek and Latin Epigraphy (EAGLE) [1] is a federation of databases, including the Epigraphic Database Roma (EDR), the Epigraphische Datenbank Heidelberg (EDH), the Epigraphic Database Bari (EDB), and Hispania Epigraphica (HE), which together contain around 350.000 inscriptions. Another important resource is the Epigraphik-Datenbank Clauss/Slaby (EDCS) [2], which provides the most extensive coverage of Latin inscriptions (around 540.000). While these platforms offer searchable texts and metadata, and some, such as EDCS, EDH, and EDR, include images for a small fraction of inscriptions, they do not support search or analysis based on visual features.

The availability of these datasets, combined with recent advances in machine learning, has contributed to the creation of a machine-readable dataset of Latin inscriptions that also includes images [9]. Nevertheless, most works in this field remain text-centered, focusing on linguistic analysis or reconstruction of lacunae [10, 11].

Among the few available Latin datasets that include images, the CATMuS dataset [12] provides a collection of handwritten documents designed specifically for the handwritten text recognition task. For Latin inscriptions, machine learning methods have so far been applied to more limited problems, such as retrieving visually similar inscriptions [13] and segmenting the inscribed text from the surrounding support [14]. More recently, Aenea [9], the current state-of-the-art model for restoration and for the geographical and chronological attribution of Latin inscriptions, incorporates visual information, but only for the task of geographical attribution. By contrast, in the context of Chinese epigraphy, Duan et al. [15] have proposed a multimodal multitask neural model that leverages both modalities directly for the restoration of ancient ideographs. These works demonstrate that while some progress has been made in integrating computer vision into Latin epigraphy, the exploitation of visual features for non-textual, iconographic elements remains unexplored compared to text-based approaches. In the Greek domain, the Bessarion dataset [16] provides Byzantine-era Medieval inscriptions with multimodal annotations, primarily aimed at supporting text recognition tasks in epigraphy.

Computer vision has been more widely applied in related areas. For example, deep learning models have been employed for the decipherment of oracle bone inscriptions, the earliest Chinese script [17], and for their recognition [18], despite their fragmentary and eroded state. Similar techniques have also been explored in medieval manuscript segmentation [19], detection of archaeological sites from satellite images [20], and numismatics [21]. In the complementary domain of art history, fine-grained annotated datasets for the iconographic analysis of paintings already exist, offering bounding-box level annotations for a wide range of common objects and iconographic figures [22, 23]. However, within the field of epigraphy, there remains a limited availability of such data, particularly datasets that are manually annotated with high-quality labels for symbol localization.

A further challenge is taxonomy design. While comprehensive taxonomies exist for general art images, such as Iconclass [24], their focus remains on art and they fail to capture the required level of detail for our specific domain. The problem of defining a shared taxonomy is also present within epigraphy, where, as [5] notes, there is no shared taxonomy across different epigraphic sources of information, which often rely on heterogeneous and non-standardized terminology. To address these challenges, specifically the need for a high-quality dataset for symbol localization and a coherent, domain-specific taxonomy for Late Antique Latin epigraphs, we introduce *EpiSignum*EpiSignum.

3. Description of EpiSignum

EpiSignum is a collection of 199 manually annotated images of Late Antique Latin inscriptions (see Figure 1 for a sample of the data), publicly available at https://github.com/AilabUdineGit/EpiSignum-code. The dataset is the result of a collaboration between computer scientists and epigraphers. The latter contributed by defining a taxonomy of the elements found in the inscriptions, reflecting the main categories of Early Christian iconography. This taxonomy serves as the basis for annotating the images and is inspired by the interpretative tradition established in classical works of Christian epigraphy and archaeology, particularly Giovanni Battista de Rossi's Inscriptiones Christianae Urbis Romae [25].

3.1. Taxonomy

The iconographic categories selected for this work reflect a historically grounded framework for interpreting these symbolic elements and aim to provide a coherent structure for the wide formal variety found in Christian visual language. The hierarchical organization adopted in the dataset captures this variety, which is influenced by several factors, including socio-cultural context, economic means of the patron, available space on the inscription, and the commemorative or symbolic function of the object itself. At the same time, the classification of early Christian symbols presents several methodological and practical challenges. The variability of recurring motifs such as the cross, the palm, or the dove depends on region, chronology, and the hand of the engraver, and often results in significant differences in shape and proportions. This heterogeneity complicates the definition of consistent criteria for the classification, hindering the integration of inscriptions into structured

collections. To this difficulty is added the frequent erosion or fragmentation of epigraphic supports, due to environmental factors, reuse, or inadequate conservation, which can render symbols only partially visible or barely legible. In such cases, interpretation has traditionally relied on the individual expertise of the scholars, introducing a high degree of subjectivity [26]. Moreover, existing digital corpora, while often rich in textual metadata and geographic information, rarely provide systematic access to the visual dimension, leaving iconographic elements underrepresented and poorly indexed. The hierarchical taxonomy adopted in this work is designed to address these issues by offering a structured framework that can accommodate formal variability and support a more coherent organization of the material. The taxonomy is organized as follows:

0. Figura humana (vir, mulier, puer, puella, adulescens)
    - orans (utraque manu, dextra manu, sinistra manu)
    - loquens (dextra manu, sinistra manu)
    - aliquid manu tenere
    - lavorans
1. Arbores et herbae (arbor, frutex, flor, palma, ramus olivae, frondes, fructus – uva, baca –, herba)
2. Architecturae pars (aedificium, columna, arcus, epistylium, velarium)
3. Elementum astrorum (astrum, sol, luna, aquae caelestes)
4. Elementum geometricum (circulus, quadratus, rectangulum)
5. Elementum naturale (collis, flumen)
6. Figura animalis
    - avis (columba vel alius)
    - ovis
    - piscis
7. Res (tabula – ansata –, volumen, corona, sertum, velum, mensa, vas, labrum, amphora, ancora, candelabrum/menorah, vestis, ornamentum, orbiculum, fibula)
8. Decussis (monogramma – cum alpha et omega –, crux latina, crux graeca, crux gammata, staurogramma)

3.2. Dataset

The images included in the dataset were gathered by epigraphers and scholars from the University of Udine and Roma Tre University, and they originate from a variety of places and chronological contexts. As a curated collection rather than a standardized capture effort, the corpus exhibits substantial variation in quality and format, with instances ranging from high-resolution black-and-white scans to color photographs and lower-resolution images. Rather than a limitation, we view this heterogeneity as an advantage: it reflects the conditions under which epigraphs are typically digitized and studied, and it provides a realistic testbed for robustness to domain shift (device, lighting, contrast, and resolution), which is essential for practical deployment in Digital Humanities workflows.

The annotation process was carried out by domain experts with specific training in the iconographic analysis of Christian inscriptions. This ensured a high degree of semantic and taxonomic accuracy, with particular care given to the identification of symbols relevant to the project. The masks were created manually: in some cases through fine-grained, pixel-level segmentation, and in others by means of bounding boxes. For the sake of uniformity, however, the public release of the dataset contains only bounding box annotations. To reduce potential bias and enhance the reliability, the work was distributed among multiple annotators, who also served as double-checkers.

Each image in the dataset is paired with a text file following the YOLO annotation format.[1] Every line of the text file corresponds to a single annotated instance, consisting of the class identifier (drilled down to the maximal applicable taxonomy depth) followed by the normalized bounding-box coordinates $(x_{\text{center}}, y_{\text{center}}, width, height)$. An example annotation file is shown below:

[1] https://docs.ultralytics.com/datasets/detect/#ultralytics-yolo-format

Class ID	Class Name	# Images	# Instances
0	Figura humana	85	108
1	Arbores et herbae	94	184
2	Architecturae pars	20	38
3	Elementum astrorum	6	15
4	Elementum geometricum	6	6
5	Elementum naturale	4	5
6	Figura animalis	83	146
7	Res	24	38
8	Decussis	134	200

Table 1
Distribution of the dataset's classes, ordered by taxonomy ID.

```
Figura_humana:puer>orans:dextra_manu 0.666667 0.597458 0.046205 0.059322
Figura_humana                        0.733498 0.739936 0.315182 0.515890
Figura_animalis>avis                 0.605611 0.530191 0.115512 0.106991
Elementum_naturale:collis           0.913366 0.461864 0.110561 0.209746
```

Note how labels encode a path through the taxonomy. Hierarchical levels are concatenated with '>', and within any level an optional subtype can be appended after ':'. Thus, for instance, `Figura_humana:puer>orans:dextra_manu` denotes the parent category *Figura humana* refined to *puer orans* specified as *dextra manu*. When no subtype applies at a given level, only the level name is used (e.g., `Figura_animalis>avis`, where just a "general" *avis* is specified).

To provide an overview of the dataset composition, Table 1 reports, for each highest-level class in the taxonomy, both the number of images containing at least one instance of that class and the total number of annotated instances across the corpus. Complementing this, Figure 3 provides a visual summary of the distribution, illustrating the frequency of each class across the dataset. In addition, Figure 4 shows the number of distinct classes represented within individual images.

The dataset exhibits a significant class imbalance. Categories such as *Decussis* and *Arbores et herbae* are highly represented, whereas *Elementum naturalis* and *Elementum geometricum* occur only rarely. In addition to this frequency imbalance, the dataset also shows variation in intra-class diversity. *Decussis* displays relatively standardized forms, since crosses tend to follow well-defined patterns. By contrast, *Arbores et herbae*, *Res*, *Architecturae pars*, and *Elementum astrorum* encompass a much wider range of visual appearances, with motifs, objects, and shapes that differ considerably from one another (see Figure 2).

Figure 2: Examples of different symbol categories. Each column represents a distinct class, with four different examples shown for each. From left to right, the classes are: *Figura humana*, *Arbores et herbae*, *Architecturae pars*, *Elementum astrorum*, *Elementum geometricum*, *Elementum naturale*, *Figura animalis*, *Res*, and *Decussis*.

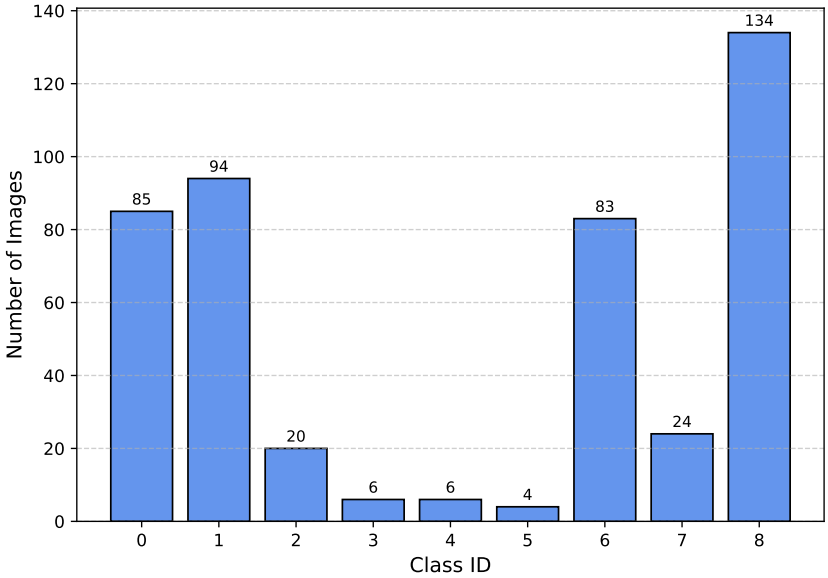

Figure 3: Number of images containing at least one instance of each class.

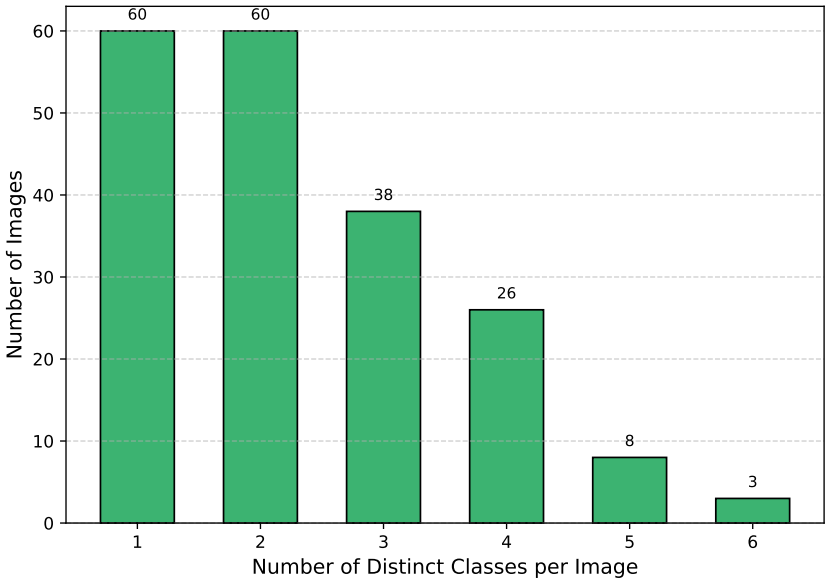

Figure 4: Number of images as a function of how many distinct classes they contain (class diversity).

4. Baseline

To analyze the dataset and establish a baseline performance reference for future work, we used YOLOv8 [27], a state-of-the-art deep learning model for object detection and classification. YOLOv8 was chosen due to its strong performance across a wide range of detection tasks [28] and its suitability for relatively small datasets [29]. For the experimental tasks, we considered the highest level of the taxonomy, neglecting subtypes, while the only preprocessing applied to the dataset was resizing all images to 960 pixels on the longer side and converting them to RGB mode if they were not already; no other (e.g., brightness, contrast) corrections were applied to the images. In addition to this supervised, closed-set baseline, we also evaluated YOLO-World [7], an open-vocabulary model that can detect objects based on arbitrary text prompts rather than a fixed set of training classes, to test the dataset's

potential for zero-shot detection tasks. To ensure reproducibility, all our code is released alongside the dataset.

Given the limited number of available examples and the unbalanced distribution of symbols across the dataset (Table 1), we adopted the Leave-One-Out Cross-Validation (LOOCV) protocol. In this setup, for each split, the model is trained on all images except one, which is used for validation. This process is repeated for every image in the dataset, ensuring that each sample serves as a validation instance exactly once. LOOCV was chosen because a fixed train/validation split could not guarantee a fair representation of all symbols, whereas this protocol ensures that every instance contributes both to training and validation.

All reported metrics correspond to the mean and standard deviation across LOOCV runs. In other words, each image produces a set of evaluation scores when used as the validation sample, and the results are then summarized for each iconographic class over all iterations.

4.1. Evaluation Metrics

To evaluate the performance of the model, we report standard metrics for the object detection task: *precision* (P) and *recall* (R), defined as follows:

$$P = \frac{TP}{TP + FP}, \quad R = \frac{TP}{TP + FN}$$

where TP, FP, and FN denote true positives, false positives, and false negatives, respectively.

For each class c, we compute the *Average Precision* (AP) at different *Intersection over Union* (IoU) thresholds. The IoU is defined as the ratio between the area of overlap and the area of union between a predicted bounding box and the corresponding ground-truth box surrounding an object. A detection is considered correct if its IoU with a ground-truth object exceeds the chosen threshold. Specifically, given a class c, we report:

- **AP50**: $AP@0.5$, i.e., AP at IoU = 0.5;
- **AP50-95**: $AP@[0.5 : 0.95]$, i.e., AP averaged over IoU thresholds from 0.5 to 0.95 with step 0.05, following the COCO evaluation protocol [30].

4.2. Zero-Shot Evaluation

As a first step, we evaluated YOLOv8 and YOLO-World in zero-shot mode, without fine-tuning. The goal was to test whether a pretrained detector could generalize directly to our iconographic classes. This choice was motivated by findings in Historical Document Analysis, where not only is data scarcity a common issue, but the size of characters is also comparable to the Late Antique Christian symbols we aim to classify, and where pretrained models have been shown to support the task [31]. However, as shown in Table 2, the zero-shot performance for both models was negligible, with metrics remaining close to zero for almost all categories. The exceptions were minimal: YOLOv8 showed modest (albeit still low) performance on 'Decussis' (Class 8), while YOLO-World showed a similar slight capability on 'Figura animalis' (Class 6) and 'Res' (Class 7). These results could be justified by the fact that the pre-trained YOLOv8 model is trained on the classes in the COCO dataset [30], which contains general everyday object categories such as persons, animals, vehicles, and household items. Although 'person' in COCO bears some affinity to our 'Figura humana' class, which may account for the modest (albeit still low) performance observed for that category, no COCO class corresponds to symbolic or stylized motifs such as crosses. Similarly, YOLO-World, despite being an open-vocabulary model designed for generalization, also proved unable to recognize these specialized classes, showing near-zero performance on all other categories. Such a class mismatch, combined with the high formal variability of early Christian symbols, discussed in Section 3, where even recurring motifs appear in markedly different shapes, proportions, or states of preservation, is not captured by generic pretrained features. In addition, many of the carved symbols in our dataset are small in scale, which represents a further difficulty for YOLO models [32].

ID	Class Name	Model	Precision	Recall	AP50	AP50-95
0	Figura humana	YOLOv8	0.100 ± 0.240	0.175 ± 0.366	0.162 ± 0.350	0.094 ± 0.224
		YOLO-World	0.160 ± 0.337	0.165 ± 0.346	0.125 ± 0.278	0.064 ± 0.149
1	Arbores et herbae	YOLOv8	0.052 ± 0.210	0.023 ± 0.125	0.025 ± 0.134	0.015 ± 0.083
		YOLO-World	0.032 ± 0.177	0.027 ± 0.153	0.027 ± 0.155	0.016 ± 0.097
2	Architecturae pars	YOLOv8	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000
		YOLO-World	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000
3	Elementum astrorum	YOLOv8	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000
		YOLO-World	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000
4	Elementum geometricum	YOLOv8	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000
		YOLO-World	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000
5	Elementum naturale	YOLOv8	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000
		YOLO-World	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000
6	Figura animalis	YOLOv8	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000
		YOLO-World	0.196 ± 0.378	0.191 ± 0.364	0.163 ± 0.329	0.088 ± 0.207
7	Res	YOLOv8	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000
		YOLO-World	0.035 ± 0.120	0.063 ± 0.224	0.028 ± 0.107	0.015 ± 0.071
8	Decussis	YOLOv8	0.015 ± 0.121	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000
		YOLO-World	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000

Table 2
Comparison of zero-shot YOLOv8 and YOLO-World results across iconographic classes. Metrics are reported as mean and standard deviation.

Class ID	Class Name	Precision	Recall	AP50	AP50-95
0	Figura humana	0.739 ± 0.233	0.897 ± 0.277	0.890 ± 0.273	0.735 ± 0.293
1	Arbores et herbae	0.793 ± 0.299	0.557 ± 0.412	0.603 ± 0.385	0.404 ± 0.325
2	Architecturae pars	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000
3	Elementum astrorum	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000
4	Elementum geometricum	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000
5	Elementum naturale	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000
6	Figura animalis	0.833 ± 0.183	0.837 ± 0.284	0.857 ± 0.259	0.637 ± 0.252
7	Res	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000
8	Decussis	0.787 ± 0.245	0.859 ± 0.311	0.856 ± 0.300	0.632 ± 0.282

Table 3
Fine-tuned (without data augmentation) YOLOv8 results across iconographic classes. Metrics are reported as mean and standard deviation.

4.3. Fine-Tuning without Data Augmentation

Next, we fine-tuned YOLOv8 on our dataset under the same LOOCV protocol. Each model was trained for 100 epochs with an early stopping patience of 20. The initial learning rate was set to 0.01. All training runs used the same hyperparameters for consistency, including batch size, learning rate, and number of epochs. No data augmentation was applied in this experiment.

As expected, Table 3 shows a clear divide between frequent, visually standardized classes and rare, heterogeneous ones. Classes such as *Decussis*, *Figura humana*, and *Figura animals* achieve relatively high performance, with precision and recall above 0.8 and $AP50$ scores exceeding 0.90 in the best cases. This is consistent with their higher representation in the dataset and, in the case of *Decussis*, the limited intra-class variability due to the standardized form of crosses.

In contrast, classes with fewer instances and greater internal diversity, such as *Res*, *Architecturae pars*, and *Elementum astrorum*, show no successful detections, with all metrics at zero. Even intermediate

Class ID	Class Name	Precision	Recall	AP50	AP50-95
0	Figura humana	0.800 ± 0.220	0.912 ± 0.240	0.899 ± 0.247	0.769 ± 0.271
1	Arbores et herbae	0.822 ± 0.266	0.651 ± 0.390	0.685 ± 0.375	0.459 ± 0.315
2	Architecturae pars	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000
3	Elementum astrorum	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000
4	Elementum geometricum	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000
5	Elementum naturale	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000
6	Figura animalis	0.858 ± 0.175	0.870 ± 0.281	0.898 ± 0.247	0.643 ± 0.255
7	Res	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000
8	Decussis	0.842 ± 0.251	0.854 ± 0.312	0.872 ± 0.293	0.662 ± 0.280

Table 4
Fine-tuned (with data augmentation) YOLOv8 results across iconographic classes. Metrics are reported as mean and standard deviation.

categories such as *Arbores et herbae* reach lower and more unstable scores, reflecting the wide variety of plant motifs and their less standardized visual forms.

4.4. Fine-Tuning with Data Augmentation

Finally, data augmentation was applied to the training images to improve generalization and to mitigate overfitting. Augmentation increases not only the amount of training data but also its diversity, which is crucial given the limited size and variability of the dataset. For each image in the training set, five synthetic images are generated by simultaneously applying a combination of all the following augmentation techniques (see the code provided alongside the dataset for additional details):

- *Geometric transformations:* random rotations, rescaling, cropping, and horizontal/vertical flips, to simulate variability in the positioning and orientation of inscriptions.
- *Photometric transformations:* adjustments of brightness and contrast, as well as the introduction of noise or blur, to reproduce real-world conditions of poor lighting or image quality.
- *Domain-specific distortions:* elastic deformations and grid distortions, designed to mimic microfractures and surface irregularities commonly found in archaeological artifacts.

Table 4 shows that augmentation yields consistent improvements over the no-augmentation baseline (Table 3), particularly for intermediate-frequency classes such as *Arbores et herbae*. Frequent and standardized classes like *Figura humana* and *Decussis* already perform well without augmentation, but still benefit slightly in terms of stability. Rare categories, however, continue to show near-zero performances across all metrics. This suggests that augmentation helps regularize the model but cannot fully compensate for severe class imbalance and high intra-class variability.

4.5. Discussion

Comparing Tables 2, 3, and 4, a clear progression can be observed. Zero-shot fails entirely, highlighting the domain specificity of the dataset. Fine-tuning without data augmentation produces good results on frequent and relatively standardized classes such as *Figura humana* and *Decussis*, but struggles with rarer, more heterogeneous ones. With augmentation, performance improves further, particularly for intermediate categories like *Arbores er herbae*, where synthetic variability helps stabilize the training. Nonetheless, rare categories remain largely undetected, reflecting the combined effect of class imbalance and intra-class variability.

Overall, these results confirm that the dataset poses a significant challenge for object detection. High performance can be achieved on abundant and standardized classes, while rare and heterogeneous ones remain very difficult to capture.

This underscores the rationale for proposing the dataset: by exposing both the strengths and limitations of standard architectures like YOLOv8 when applied to epigraphic material, it provides a

benchmark for systematic evaluation and a foundation for developing domain-adapted methods capable of generalizing across underrepresented iconographic classes.

5. Conclusion and future work

This paper introduced *EpiSignum*, a taxonomy and annotated dataset for detecting symbols, that is, non-textual visual elements, in Late Antique Latin inscriptions. Our aim is to address the scarcity of high-quality labeled data and the lack of a consistent framework for interpreting these motifs. Experiments with a strong general-purpose detector, YOLOv8, reveal considerable limitations, with performance that remains poor for this specialized task. Data scarcity and a long-tailed class distribution also constrain the effectiveness of straightforward fine-tuning, pointing to the need for tailored approaches. We proposed a data-augmentation pipeline that yields measurable gains, while better results might be achievable by employing more realistic, structure-aware data synthesis. We believe that *EpiSignum* can serve as a reproducible benchmark and a catalyst for progress in epigraphic symbol analysis.

As future work, we plan to expand the dataset to increase coverage and enable finer-grained classification within the taxonomy; to move from bounding-box detection to instance segmentation in order to capture symbol shape; and to incorporate textual transcriptions and metadata so that multimodal models can jointly exploit visual and textual cues. Beyond these steps, we see promise in self-supervised and domain-specific pretraining on large unlabeled epigraphic corpora; and in human-in-the-loop active learning with epigraphers to target uncertain or rare cases.

Acknowledgments

This work was supported by the Department Strategic Plan (DSP) of the University of Udine—Interdepartment Projects: Artificial Intelligence, and the DIUM-DMIF.

Data and Code Availability

The dataset used in this study is available at https://github.com/AI4CH-UniUD/EpiSignum. The code supporting the findings of this work is available at https://github.com/AilabUdineGit/EpiSignum-code.

Declaration on Generative AI

During the preparation of this work, the author(s) used ChatGPT-5 and Grammarly in order to: Grammar and spelling check. After using these tool(s)/service(s), the author(s) reviewed and edited the content as needed and take(s) full responsibility for the publication's content

References

[1] Eagle: Europeana network of ancient greek and latin epigraphy, https://www.eagle-network.eu/, 2013. Best Practice Network co-funded by the European Commission's ICT Policy Support Programme; project duration: 1 April 2013–31 March 2016.

[2] Epigraphik-datenbank clauss/slaby, https://db.edcs.eu/epigr/, 1985. Online database of Latin inscriptions; founded by Manfred Clauss and Wolfgang A. Slaby in the mid-1980s, based at the Katholische Universität Eichstätt-Ingolstadt.

[3] Y. Assael, T. Sommerschield, J. Prag, Restoring ancient text using deep learning: a case study on greek epigraphy, arXiv preprint arXiv:1910.06262 (2019).

[4] Y. Assael, T. Sommerschield, B. Shillingford, M. Bordbar, J. Pavlopoulos, M. Chatzipanagiotou, I. Androutsopoulos, J. Prag, N. De Freitas, Restoring and attributing ancient texts using deep neural networks, Nature 603 (2022) 280–283.

[5] V. Kase, P. Heřmánková, A. Sobotková, Classifying latin inscriptions of the roman empire: A machine-learning approach, in: Computational Humanities Research 2021, 2021, pp. 123–135.
[6] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, You only look once: Unified, real-time object detection, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.
[7] T. Cheng, L. Song, Y. Ge, W. Liu, X. Wang, Y. Shan, Yolo-world: Real-time open-vocabulary object detection, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2024, pp. 16901–16911.
[8] A. Buonopane, Manuale di epigrafia latina, Beni culturali, Carocci, Roma, 2009. Tex.lccn: 2009478450.
[9] Y. Assael, T. Sommerschield, A. Cooley, B. Shillingford, J. Pavlopoulos, P. Suresh, B. Herms, J. Grayston, B. Maynard, N. Dietrich, et al., Contextualizing ancient texts with generative neural networks, Nature (2025) 1–7.
[10] T. Clérice, Detecting sexual content at the sentence level in first millennium latin texts, arXiv preprint arXiv:2309.14974 (2023).
[11] A. Brunello, E. Colombi, A. Locaputo, S. Magnani, N. Saccomanno, G. Serra, et al., Usage of language model for the filling of lacunae in ancient latin inscriptions: A case study, in: Proceedings of the 2nd Workshop on Artificial Intelligence for Cultural Heritage (IAI4CH 2023) co-located with the 22nd International Conference of the Italian Association for Artificial Intelligence (AIxIA 2023), Roma, Italy, volume 3536, 2023, pp. 113–125.
[12] T. Clérice, A. Pinche, M. Vlachou-Efstathiou, A. Chagué, J.-B. Camps, M. G. Levenson, O. Brisville-Fertin, F. Boschetti, F. Fischer, M. Gervers, et al., Catmus medieval: A multilingual large-scale cross-century dataset in latin script for handwritten text recognition and beyond, in: International Conference on Document Analysis and Recognition, Springer, 2024, pp. 174–194.
[13] G. Amato, F. Falchi, L. Vadicamo, Visual recognition of ancient inscriptions using convolutional neural network and fisher vector, Journal on Computing and Cultural Heritage (JOCCH) 9 (2016) 1–24.
[14] C. Tupman, D. Kangin, J. Christmas, Reconsidering the roman workshop: using computer vision to analyse the making of ancient inscriptions, Umanistica Digitale (2021) 461–473.
[15] S. Duan, J. Wang, Q. Su, Restoring ancient ideograph: a multimodal multitask neural network approach, arXiv preprint arXiv:2403.06682 (2024).
[16] G. Sfikas, P. Dimitrakopoulos, G. Retsinas, C. Nikou, P. Kitsiou, Bessarion: Medieval greek inscriptions on a challenging dataset for vision and nlp tasks, in: International Workshop on Document Analysis Systems, Springer, 2024, pp. 393–407.
[17] H. Guan, H. Yang, X. Wang, S. Han, Y. Liu, L. Jin, X. Bai, Y. Liu, Deciphering oracle bone language with diffusion models, arXiv preprint arXiv:2406.00684 (2024).
[18] M. Liu, G. Liu, Y. Liu, Q. Jiao, Oracle bone inscriptions recognition based on deep convolutional neural network, Journal of image and graphics 8 (2020) 114–119.
[19] A. De Nardin, S. Zottin, C. Piciarelli, E. Colombi, G. L. Foresti, Few-shot pixel-precise document layout segmentation via dynamic instance generation and local thresholding, International Journal of Neural Systems 33 (2023) 2350052.
[20] G. Caspari, P. Crespo, Convolutional neural networks for archaeological site detection–finding "princely" tombs, Journal of Archaeological Science 110 (2019) 104998.
[21] R. Cabral, M. De Iorio, A. Harris, From coin to data: the impact of object detection on digital numismatics, Digital Scholarship in the Humanities (2025) fqaf046.
[22] G. Strezoski, M. Worring, Omniart: a large-scale artistic benchmark, ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM) 14 (2018) 1–21.
[23] N. Gonthier, Y. Gousseau, S. Ladjal, O. Bonfait, Weakly supervised object detection in artworks, in: Proceedings of the European conference on computer vision (ECCV) workshops, 2018, pp. 0–0.
[24] L. D. Couprie, Iconclass: an iconographic classification system, Art libraries journal 8 (1983) 32–49.
[25] G. B. De Rossi, Inscriptiones Christianae urbis Romae septimo saeculo antiquiores, volume 2, Ex officina libraria Pontificia, 1888.

[26] A. E. Cooley, epigraphy, latin, 2025. URL: https://oxfordre.com/classics/view/10.1093/acrefore/9780199381135.001.0001/acrefore-9780199381135-e-2451. doi:10.1093/acrefore/9780199381135.013.2451.

[27] G. Jocher, J. Qiu, A. Chaurasia, Ultralytics YOLO, 2023. URL: https://github.com/ultralytics/ultralytics.

[28] Z. Hua, K. Aranganadin, C.-C. Yeh, X. Hai, C.-Y. Huang, T.-C. Leung, H.-Y. Hsu, Y.-C. Lan, M.-C. Lin, A benchmark review of yolo algorithm developments for object detection, IEEE Access (2025).

[29] S. Zhang, Y. Yang, L. Tu, T. Fu, S. Chen, F. Cen, S. Yang, Q. Zhao, Z. Gao, T. He, Comparison of yolo-based sorghum spike identification detection models and monitoring at the flowering stage, Plant Methods 21 (2025) 20.

[30] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, C. L. Zitnick, Microsoft coco: Common objects in context, in: European conference on computer vision, Springer, 2014, pp. 740–755.

[31] L. Studer, M. Alberti, V. Pondenkandath, P. Goktepe, T. Kolonko, A. Fischer, M. Liwicki, R. Ingold, A comprehensive study of imagenet pre-training for historical document image analysis, in: 2019 International Conference on Document Analysis and Recognition (ICDAR), IEEE, 2019, pp. 720–725.

[32] W. Saenprasert, E. E. Tun, A. Hajian, W. Ruangsang, S. Aramvith, Yolo for small objects in aerial imagery: A performance evaluation, in: 2024 21st International Joint Conference on Computer Science and Software Engineering (JCSSE), IEEE, 2024, pp. 720–727.