Classifying Name-Date and Year Figures in Mixtec Codices

Girish Salunke, Christopher Driggers-Ellis[1], and Christan Grant[1]

[1] Computer & Information Science & Engineering, University of Florida, Gainesville, FL, USA

Abstract

The Mixtec codices are precolonial and early colonial Mesoamerican manuscripts that use a semasiographic writing system to record historical, genealogical, and cosmological information. Decoding these manuscripts is uniquely challenging due to their non-linear structure and symbolic complexity. In this work, we introduce a novel application of Vision Transformers (ViTs) to classify key elements of Mixtec writing. Our classification pipeline involves a type classifier to distinguish between year and name-date symbols, followed by a symbol classifier to identify calendrical signs, and finally, a numerical bead counter. This pipeline tackles the challenge of high intra-class variability in hand-drawn symbols and separates the complex, often overlapping symbolic glyphs from their numerical components. Our results show a surprising dichotomy: ViTs achieve an F1 score greater than 0.9 in symbol classification but struggle with the counting task, where the F1 score is about 0.22. This contrast highlights a core architectural trade-off in Vision Transformers: their global attention mechanism helps with holistic pattern recognition, but it hinders fine-grained spatial localization. This insight clarifies both the potential and limitations of using ViTs to decode semasiographic texts, which can help guide more targeted applications in cultural heritage preservation. Our code is available here: `https://github.com/ufdatastudio/mixtec-namedate-classifiers`

Keywords: Mixtec Codices, Vision Transformers, Semasiography, Cultural Heritage, Image Classification, Binary Classification, Mesoamerican Studies, Mesoamerican Manuscripts, Pictographic Manuscripts, Semasiographic Manuscripts, Pictographic Writing

1 Introduction

The Mixtec codices are manuscripts produced by the Mixtec civilization of pre-colonial and early colonial Mesoamerica. These codices use a semasiographic writing system—conveying meaning through images rather than phonetic or logographic symbols. Rich with imagery representing historical events, deities, rulers, and calendrical information, the codices function as complex narrative documents that communicate cultural knowledge independently of phonetic representation. Analyzing these narratives is fundamental to preserving the cultural memory of the Mixtec people and provides a vital indigenous perspective on Mesoamerican history. Understanding these intricate systems is crucial not only for the Mixtec community but also for broader Mesoamerican studies, offering unique insights into pre-Columbian thought, history, and social structures that are often absent from colonial records.

Deciphering Mixtec writing is challenging due to its semasiographic nature, limited annotated data, and nonlinear reading order, often guided by visual cues within the codices. Recent advances in computer vision have proven effective for historical document analysis; however, existing methods are largely designed for phonetic scripts. The decoding of semasiographic systems remains an

Girish Salunke, Christopher Driggers-Ellis, and Christan Grant. "Classifying Name-Date and Year Figures in Mixtec Codices." In: *Computational Humanities Research 2025*, ed. by Taylor Arnold, Margherita Fantoli, and Ruben Ros. Vol. 3. Anthology of Computers and the Humanities. 2025, 1178–1186. https://doi.org/10.63744/eNge3Gj1LSHb.

underexplored area, particularly in low-resource contexts where symbol sets are small but semantically dense [7].

In this work, we explore the use of Vision Transformers (ViTs)—a class of transformer-based models adapted for image understanding—to decode key elements of Mixtec codices. Our goal is to classify name-dates and year symbols, which are central to interpreting the narratives in these manuscripts. This paper outlines our contributions, including: (1) proposing a novel ViT-based pipeline for semasiographic symbol classification, (2) demonstrating the efficacy of ViTs for distinguishing complex glyphs in low-resource settings, and (3) highlighting the architectural limitations of ViTs for fine-grained counting tasks, which is critical for future research in computational cultural heritage.

1.1 Background

The Mixtec calendar system, adapted from the broader Mesoamerican tradition, is built on a 260-day Sacred Calendar (Tonalpohualli) and a 365-day Solar Calendar (Xiuhpohualli) [16]. Individuals are named for days in the Sacred Calendar, with the associated date typically corresponding to their date of birth. These name-dates are composed of a number (1–13) and one of 20 calendrical day signs (e.g., "8 Deer," "4 Reed") [2; 16]. These name-dates serve as proper names for historical figures, gods, or ancestors and often recur across multiple codices. Similarly, years are marked using a combination of a number (1–13) and one of four year signs (Rabbit, Reed, House, Flint), forming a 52-year Calendar Round used for historical chronology [2; 16].

Visually, name-dates and year symbols share structural similarities, typically consisting of a numeric prefix depicted as a series of small "beads"—paired with a pictorial glyph. However, their *semantic roles* differ: name-dates identify individuals (typically based on the calendar day of birth), while year symbols serve as chronological anchors for historical events. The presence of the *year AO glyph* is used to disambiguate these cases in several codices, including the Zouche-Nuttall.[1] Figure 1 visually illustrates this distinction, with name-dates highlighted using red squares and year symbols marked in green. Importantly, the term *name-date* also applies to instances where a specific calendar day (e.g., "8 Deer") is appended to a year (e.g., "3 Flint") to record the exact date of a historical event. In such cases, the name-date functions not as a personal identifier but as a temporal marker within the 52-year calendar round. Both types of symbols draw from the same visual inventory of calendrical signs and follow a similar structural format. This shared design creates considerable visual ambiguity, especially when the presence of the AO glyph is faint, damaged, or omitted. As a result, distinguishing between name-dates and year symbols is not only essential for determining *who is depicted* and *when an event occurred*, but also a non-trivial classification challenge that requires careful symbolic analysis [2; 16].

To address this challenge, we propose a image classification sequence using ViTs. Figure 3 illustrates this using an example:

- The **name-date/year classifier** distinguishes between name-dates and year symbols;
- The **symbol classifier** identifies the calendrical glyph;
- The **bead counter** module extracts the numerical prefix.

Our approach leverages transfer learning and data augmentation to address the limited size of the annotated corpus. This study contributes to the growing body of research at the intersection of machine learning and cultural heritage, demonstrating how transformer-based architectures can assist in interpreting symbolic, non-phonetic writing [6; 9; 14].

[1] The "AO" year sign, also known as the *annus obitus* or "year-of" glyph, is a conventional Mixtec symbol that marks a calendrical expression as a reference to a year rather than a personal name. It often appears as a stylized circle or cartouche enclosing the year symbol [2; 16].

Figure 1: Red squares highlight name-dates, which identify individuals through a combination of a day sign and number. The green square highlights a year symbol, which serves as a chronological marker within the narrative. Provided courtesy of the British Museum.

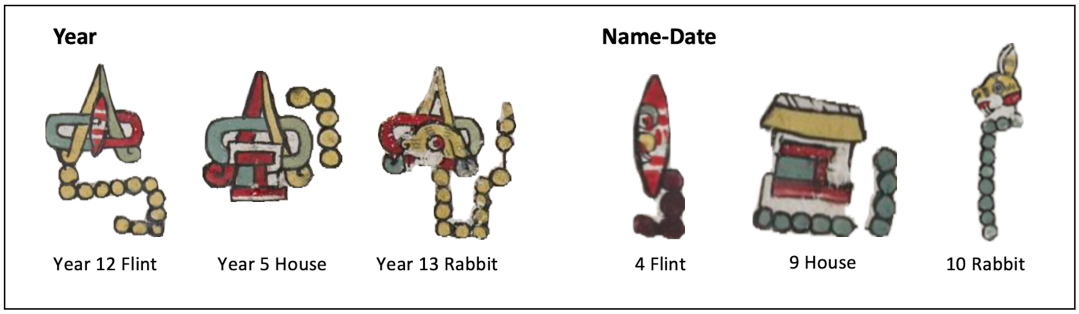

Figure 2: Visual comparison of a **name-date** and a **year symbol with AO marker**. Both include a numerical prefix (beads) and a glyph (e.g., Rabbit or Reed), but the year symbol includes the AO marker, distinguishing it as a chronological reference rather than a personal identifier. Cropped from original images provided courtesy of the British Museum.

2 Related Work

Machine-based interpretation of historical manuscripts has long been an interdisciplinary challenge, at the intersection of computer vision, anthropology, and linguistics. Although considerable progress has been made in the analysis of phonetic scripts using OCR and handwriting recognition [17], semasiographic systems, such as those found in Mesoamerican codices, require different approaches due to their non-linear structure and reliance on symbolic imagery.

Studies on Mesoamerican writing systems show that these codices follow strict conventions to depict years, dates, names, and individuals [2; 8]. The Mixtec codices, in particular, use calendrical and genealogical references through symbols rather than phonetic text, making visual classification crucial for decipherment. Foundational ethnographic and iconographic analyses laid the groundwork for understanding how meaning is encoded through symbolic combinations of numbers and day or year glyphs [2; 13].

Recent computational work has begun to operationalize these iconographic insights. One related work introduces a dataset of segmented figures from three Mixtec codices, Zouche-Nuttall, Selden, and Vindobonensis, and apply deep learning models (VGG-16 and ViT-B/16) to two binary classification tasks: gender and pose [15]. Their experiments demonstrate that fine-tuned Vision Transformers outperform CNNs, particularly at higher learning rates, and that attention map analyses reveal alignment between model predictions and expert heuristics (e.g., visual cues like loincloths or skirts). Our work validates the feasibility deep learning approaches to low-resource

cultural heritage datasets.

Name-dates function as identifiers usually assigned at birth according to the 260-day ritual calendar, while year symbols locate events within a 52-year cycle. Name-dates are essential for understanding who is involved and when an event occurred in the codices, bridging the gap between image analysis and historical understanding. Related work explores attributes such as pose and gender to help identify individual figures [15]. In contrast, our classifiers interpret the temporal and referential structures that form the manuscript's narrative backbone.

Beyond Mixtec codices, other work in historical document analysis has applied deep learning to ancient Egyptian hieroglyphs [1] and Mayan glyphs [3]. However, these efforts often rely on large labeled datasets or exploit textual regularities not present in Mixtec imagery. Our use of Vision Transformers demonstrates the potential of transformer-based architectures for low-resource, symbolic domains like semasiographic scripts.

Particularly, our approach leverages Vision Transformers (ViTs) to offer a promising view of global visual context—particularly in cases where symbols are semantically rich but structurally ambiguous. We decode symbolic elements in Mixtec codices by classifying their relevant semantic components including glyph type, identity, and number using ViTs fine-tuned via transfer learning (Figure 3).

3 Methodology

The visual complexity and symbolic structure of Mixtec codices require a tailored computational approach that can distinguish between semantically distinct yet visually similar glyphs. We develop a modular classification pipeline that isolates and analyzes key calendrical components, specifically name-dates and year symbols, using fine-tuned Vision Transformer (ViT) models. This methodology section outlines the architecture of our system, training strategy, and data preprocessing steps, providing a foundation for automated calendrical interpretation in semasiographic scripts.

3.1 Overview of Classification Pipeline

The input to our system is a cropped image segment from annotated Mixtec codices. The classification pipeline consists of the following components:

- **Name-Date/Year Classifier**: Distinguishes whether a given symbol represents a year or a name by detecting the AO symbol.
- **Symbol Classifier**: Identifies the glyph symbol, such as "Rabbit," "Reed," or "Flint."
- **Bead Counter**: Predicts the numerical component (1–13) typically represented as a sequence of small circular marks or "beads" accompanying the main symbol.

Our classification pipeline is designed to decode calendrical symbols found in the Mixtec codices by identifying their type, symbolic content, and associated number. The input to this system consists of cropped image segments containing annotated name-date and year symbols, taken from the dataset introduced in [15]. This data set includes high-resolution scans of three Mixtec codices: Codex Zouche-Nuttall, Codex Selden, and Codex Vindobonensis Mexicanus I, and was initially annotated for figure-level classification tasks such as gender and pose. We extend this work by labeling symbolic units corresponding to Mixtec calendrical representations.

Figure 3 illustrates the proposed classification pipeline. Images are processed by three classifiers: the Name-Date/Year Classifier, the Symbol Classifier, and the Bead Counter. The Name-Date/Year Classifier determines whether the symbol represents a name-date or a year. Both categories share a common structure: a pictographic glyph accompanied by one or more small circular marks or "beads" representing a numerical value of 1 to 13, but differ in their semantic roles. To classify an input as a year, the model must detect the presence of the AO symbol. The Symbol

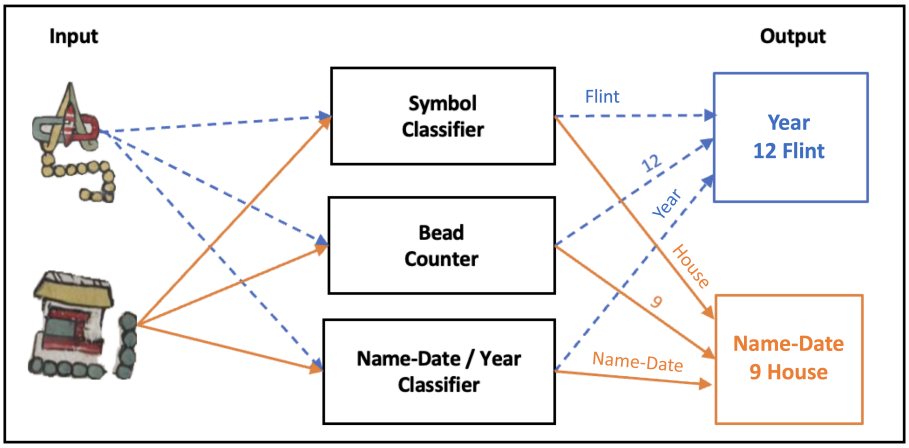

Figure 3: Proposed classification pipeline: Input images are fed into independent classifiers: a Name-Date/Year Classifier determines the symbol's type, a Symbol Classifier identifies the calendrical glyph (e.g., Reed or House), and a Bead Counter extracts the associated numerical value. The outputs are then combined to form the complete calendrical interpretation, such as "Name-Date Year 12 Reed" or "Year 9 House."

Classifier identifies the glyph itself. For year symbols, the classifier classifies the glyph as one of four canonical symbols: Rabbit, Reed, House, or Flint—associated with the 365-day solar calendar. For name-dates, the classifier selects from a set of 20 day signs used in the 260-day ritual calendar (e.g., Deer, Jaguar, Rain, Flower) [16].

The Bead Counter, which counts the number of beads depicted alongside the glyph. This number completes the calendrical representation by forming a pair with the glyph's sign (e.g., "8 Deer" or "3 Flint", see Figure 2).

We implement each component in the pipeline using a fine-tuned Vision Transformer (ViT-B/16) model [5] trained on task-specific subsets of the annotated symbol dataset.

3.2 Vision Transformer Architecture

We fine-tune a Vision Transformer (ViT-B/16) model [5] for each of the classification tasks described in our pipeline. This architecture is particularly effective in capturing both local and global visual dependencies, making it suitable for distinguishing the subtle symbolic differences present in Mixtec codices.

We initialize our ViT-B/16 models with default weights from the `torchvision.models` module. For each classification task—type identification (name-date vs year), glyph classification, and bead counting, we replace the original classification head with a task-specific linear layer. The backbone of the pre-trained ViT is frozen to preserve general visual features, while we train the classification head on our Mixtec symbol dataset.

We adopt cross-entropy loss and the Adam optimizer with a learning rate of 0.001. To enhance reproducibility, we fix random seeds and use the automated image pre-processing pipeline associated with the pre-trained ViT weights, including resizing, normalization, and color adjustments.

We perform fine-tuning using a single A100 GPU, with a batch size of 32 and up to 16 parallel data-loading workers. We trained each classifier for 10 epochs, as visual inspection of the loss curves in Figure 4 indicated plateaus in training and test loss.

Given the limited size of our dataset, we employ transfer learning to leverage the representational power of ViTs pre-trained on large-scale natural image corpora.

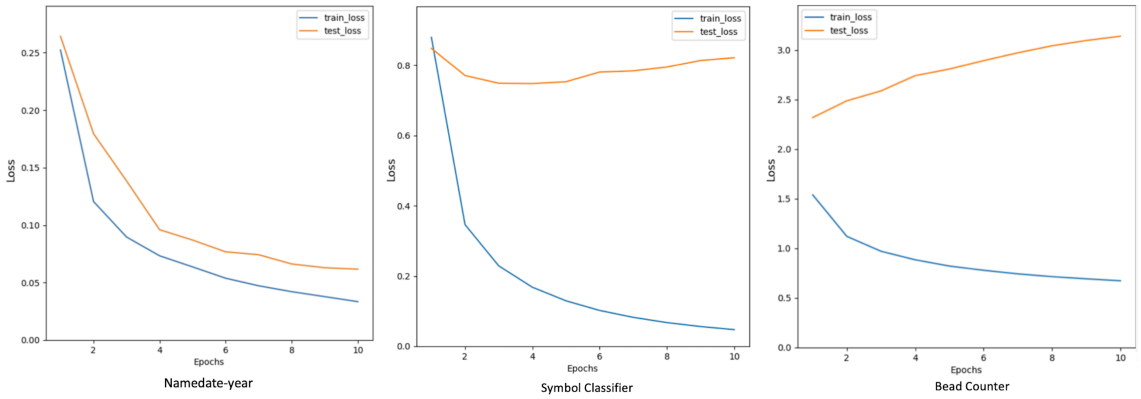

Figure 4: Loss Curves for ViT Classifiers - (a) Namedate-year, (b) Symbol Classifier, and (c) Bead Counter. Each subplot illustrates the progression of both training and testing loss over 10 epochs.

3.3 Data Augmentation

To enhance model generalization and mitigate overfitting due to limited training data, we apply four data augmentation techniques to the Mixtec symbol images. The dataset is partitioned into training and test sets in a 70:30 ratio, yielding 48,742 training images and 20,889 test images. Data augmentation is applied iteratively, with each augmentation technique performed sequentially. After each augmentation step, the augmented images are added to the existing dataset, progressively increasing the total number of images.

Random Rotation: Images are rotated at varying angles to simulate the diverse orientations in which symbols may appear across different codices.

Color Jittering: Controlled adjustments to brightness, contrast, and hue are applied to account for visual inconsistencies such as faded ink, aged parchment, and scanning artifacts.

Random Masking: Portions of the symbols are randomly occluded to encourage the model to learn discriminative features from partial inputs and improve resilience to damage or incomplete glyphs.

Oversampling of Underrepresented Classes: Classes with fewer examples (especially rarer day or year symbols) are augmented more heavily to balance the class distribution within the dataset.

These augmentations, improving the robustness of the classifiers to visual noise, partial occlusions, and layout inconsistencies commonly found in historical manuscripts.

4 Results

Table 1 shows the test results after 10 epochs of training. The *type classifier* and *symbol classifier* achieve better performance than *bead counter*. The binary *type classifier* achieves an F1 score >.90, and *symbol classifier* achieves F1 score of 0.74 despite the complexity of classifying 20 visually similar glyphs. In contrast, *bead counter* shows a >.50 gap between training and test accuracy, suggesting that Vision Transformers are ineffective for counting repetitive visual elements.

Vision Transformers (ViTs), while effective at symbolic classification, performed poorly in *bead counter* tests, with an F1 score of 0.22. The beads in the Mixtec codices are small, repetitive, and closely spaced, conditions that challenge ViTs, which lack the local spatial biases that convolutional networks naturally possess [4].

Classifier	Train Accuracy	Test Accuracy	Precision	Recall	F1
Type Classifier	0.9892	0.9795	0.9697	0.9209	0.9436
Symbol Classifier	0.9773	0.8562	0.7491	0.7318	0.7367
Bead Counter	0.7748	0.2589	0.2372	0.2201	0.2283

Table 1: Training accuracy and test accuracy, precision, recall and F1 score for each classifier after 10 epochs.

5 Discussion

Our experiments demonstrate that ViTs can be effectively fine-tuned to classify symbolic components in Mixtec codices. The strong performance of the type and symbol classifiers suggests that ViTs are capable of modeling nuanced visual distinctions between glyph categories, even under limited data conditions [5; 15].

However, ViTs performed poorly on the bead counting task, with an F1 score around 0.22. This is likely due to their patch-based processing and global attention mechanism, which makes them less suitable for tasks requiring fine-grained spatial precision [10]. Beads in Mixtec codices are small, repetitive, and closely spaced conditions that challenge ViTs, which lack the local spatial biases that convolutional networks naturally possess [4]. These results suggest that object detection or localization-based models (e.g., Faster R-CNN, YOLO) may be more appropriate for counting tasks in this domain [11; 12].

6 Future Work

The classifiers developed in this work—focused on identifying name-dates, year symbols, and their numeric prefixes serve as foundational components for building a machine interpreter capable of analyzing Mixtec codices at scale. When integrated with other visual classifiers, these modules can support more comprehensive semantic parsing of codical scenes.

Our current results are promising for symbolic classification; future work must explore more suitable models for tasks requiring fine-grained spatial reasoning, such as bead counting. Object detection or segmentation-based architectures (e.g., YOLO, Mask R-CNN) may offer greater accuracy for localizing and enumerating repeated visual units like beads. Incorporating such models would strengthen the pipeline's ability to extract structured temporal data from codices.

Finally, although our work focuses on the Mixtec codices, the underlying architecture and methodology are extensible to other low-resource semasiographic writing systems, such as Zapotec, Aztec, or Maya codices, where meaning is conveyed visually rather than phonetically. Future research will aim to generalize these classifiers across such systems.

7 Limitations

Several of the challenges we encountered reflect broader limitations previously identified in the study of Mixtec codices [15], particularly those related to dataset scope and classification methodology. While the Mixtec civilization produced a rich corpus of codices, conquest and the passage of time have left only a few high-quality manuscripts [2]. Although digitized versions of these codices support computational analysis, the limited number and stylistic variation of surviving sources constrain dataset diversity and generalizability.

Fine-tuning of ViTs from pre-trained weights introduces additional constraints. These models are initialized with parameters derived from large-scale image datasets outside the domain of Mesoamerican visual culture. As a result, they may carry latent biases that are not explicitly visible

in classification output but could influence downstream tasks. We have not yet investigated the nature or extent of these inherited biases.

Furthermore, our model does not incorporate context. Symbols are classified in isolation, despite the fact that their meaning in the codices is often compositional. This limitation is particularly relevant for name-dates, which may function either as personal identifiers or as temporal markers depending on their placement within a scene. Without considering spatial and relational context, such as proximity to figures or year glyphs, ViTs may misinterpret what role a name-date plays.

Finally, we have not integrated feedback loops with domain experts or stakeholders in the Mixtec community but we hope this work will spur these necessary collaborations.

References

[1] Barucci, Andrea, Cucci, Costanza, Franci, Massimiliano, Loschiavo, Marco, and Argenti, Fabrizio. "A Deep Learning Approach to Ancient Egyptian Hieroglyphs Classification". In: *IEEE Access 9* (2021), pp. 123438–123447. DOI: 10.1109/ACCESS.2021.3110082.

[2] Boone, Elizabeth Hill. *Stories in red and black : pictorial histories of the Aztecs and Mixtecs / Elizabeth Hill Boone.* eng. 1st ed. Austin, Texas, USA: University of Texas Press, 2000. ISBN: 0292708769.

[3] Can, Gülcan, Odobez, Jean-Marc, and Gatica-Perez, Daniel. "How to Tell Ancient Signs Apart? Recognizing and Visualizing Maya Glyphs with CNNs". In: *J. Comput. Cult. Herit. 11*, no. 4 (Dec. 2018), pp. 1–25. ISSN: 1556-4673. DOI: 10.1145/3230670.

[4] d'Ascoli, Stéphane, Touvron, Hugo, Leavitt, Matthew L, Morcos, Ari S, Biroli, Giulio, and Sagun, Levent. "ConViT: improving vision transformers with soft convolutional inductive biases*". In: *Journal of Statistical Mechanics: Theory and Experiment 2022*, no. 11 (Nov. 2022), p. 114005. ISSN: 1742-5468. DOI: 10.1088/1742-5468/ac9830.

[5] Dosovitskiy, Alexey. "An image is worth 16x16 words: Transformers for image recognition at scale". In: *arXiv preprint arXiv:2010.11929*. 2020.

[6] Fiorucci, Marco, Khoroshiltseva, Marina, Pontil, Massimiliano, Traviglia, Arianna, Del Bue, Alessio, and James, Stuart. "Machine Learning for Cultural Heritage: A Survey". In: *Pattern Recognition Letters 133* (2020), pp. 102–108. ISSN: 0167-8655. DOI: 10.1016/j.patrec.2020.02.017.

[7] Iglesia, Martin de la, Diehr, Franziska, Sikora, Uwe, Gronemeyer, Sven, Behnert-Brodhun, Maximilian, Prager, Christian, and Grube, Nikolai. "The Code of Maya Kings and Queens: Encoding and Markup of Maya Hieroglyphic Writing". In: *Journal of the Text Encoding Initiative 14* (2021).

[8] Jansen, Maarten. *The Mesoamerican Codices*. British Museum, Great Russell Street, London WC1B 3DG, United Kingdom: British Museum Publications, 1988.

[9] Muther, Ryan, Barber, Mathew, and Smith, David. "Querying the Past: Automatic Source Attribution with Language Models". In: *Proceedings of the Computational Humanities Research Conference 2023*. Paris, France, Dec. 2023, pp. 344–355.

[10] Naseer, Muzammal et al. "Intriguing properties of vision transformers". In: *arXiv preprint arXiv:2105.10497* (2021).

[11] Redmon, Joseph et al. "You Only Look Once: Unified, Real-Time Object Detection". In: *CVPR*. Las Vegas, Nevada, USA, 2016.

[12] Ren, Shaoqing et al. "Faster R-CNN: Towards real-time object detection with region proposal networks". In: *Advances in Neural Information Processing Systems*. Montréal, Canada, 2015.

[13] Smith, Mary Elizabeth. *Picture Writing from Ancient Southern Mexico: Mixtec Place Signs and Maps*. Civilization of the American Indian Series. 2800 Venture Drive Norman, Oklahoma, USA 73069: University of Oklahoma Press, 1973.

[14] Ströbel, Phillip Benjamin, Guo, Zejie, Karagöz, Ülkü, Willi, Eva Maria, and Maier, Felix K. "Bringing Rome to life: evaluating historical image generation". In: *Proceedings of the Computational Humanities Research Conference 2024*. CEUR Workshop Proceedings. Aarhus, Denmark: CEUR-WS, Dec. 2024, pp. 113–126. DOI: 10.5167/uzh-265462.

[15] Webber, Alexander, Sayers, Zachary, Wu, Amy, Thorner, Elizabeth, Witter, Justin, Ayoubi, Gabriel, and Grant, Christan. "Analyzing Finetuned Vision Models for Mixtec Codex Interpretation". In: *Proceedings of the 4th Workshop on Natural Language Processing for Indigenous Languages of the Americas (AmericasNLP 2024)*, ed. by Manuel Mager, Abteen Ebrahimi, Shruti Rijhwani, Arturo Oncevay, Luis Chiruzzo, Robert Pugh, and Katharina von der Wense. Mexico City, Mexico: Association for Computational Linguistics, June 2024, pp. 42–49. DOI: 10.18653/v1/2024.americasnlp-1.6.

[16] Williams, Robert Lloyd. *The complete codex Zouche-Nuttall : Mixtec lineage histories and political biographies / by Robert Lloyd Williams ; foreword by Rex Koontz*. eng. 1st ed. Linda Schele series in Maya and pre-Columbian studies. Austin, Texas, USA: University of Texas Press, 2013. ISBN: 9780292744387.

[17] Yadav, Divakar, Sanchez-Cuadrado, Sonia, and Morato, Jorge. "Optical Character Recognition for Hindi Language Using a Neural-network Approach". In: *Journal of Information Processing Systems 9*, no. 1 (Jan. 2013), pp. 117–140. DOI: 10.3745/JIPS.2013.9.1.117.