Computational Map Semiotics: Detecting and Encoding Cartographic Signs in Large Heterogeneous Historical Corpora

Rémi Petitpierre
EPFL, Swiss Federal Institute of Technology in Lausanne.
Digital Humanities Laboratory (DHLAB), School of Computer and Communication Sciences.
Urban Sociology Laboratory (LASUR), School of Architecture, Civil and Environmental Engineering.
CM 1 468, Station 10, 1015 Lausanne, Switzerland.
remi.petitpierre@epfl.ch

Abstract

This work introduces methods for decoding and encoding cartographic signs at scale. First, it uses class-agnostic object detection to extract 63 million signs from 99,715 historical maps. It then adapts vision foundation models to produce expressive, interpretable visual embeddings while suppressing material confounders, supporting comparative research in map semiotics and historical geography.

1 Introduction

Symbols and icons in archival maps represent an expansive source of evidence for historical geography and semiotics. However, they have never been targeted at scale, and the few existing works never engaged with the diversity of map signs, aiming at robust, scalable, and generalizable detection and encoding (Smith et al., 2025; Sutty and Duménieu, 2024). In this work, we present a two-stage framework for detecting cartographic signs in historical maps and generating interpretable sign embeddings that capture both semantics and stylistic variations. We then demonstrate the scalability of the approach, and the potential for discovery by extracting 63 million signs from 99,715 historical maps and visualizing them. The methodological challenge consists in making point signs, defined as geographic features depicted as discrete, extensionless cartographic objects, tractable at scale, while maintaining an explicitly generic scope across collections, sign repertoires, and cartographic styles. The present research frames automated sign extraction less as a tool for exhaustive reconstruction than one for distributional analysis of map signs, enabling comparative research on historical geography and cartographic semiotics.

2 Automated detection of map signs

The first part of the study assesses the practicality of automated point sign extraction from historical maps, robust to both content and form diversity. In this perspective, CaSiDD, the Cartographic Sign Detection Dataset, is introduced as an open benchmark resource comprising 18,750 sign labels from 796 distinct map patches, curated from 20 distinct map libraries across five countries (Petitpierre and Jiang, 2025). Annotation is performed in a class-agnostic fashion rather than following a closed ontology of geographic-cartographic objects. This design choice aims to broaden the scope of detection capabilities by engaging with the diversity of existing map symbols. The operationalization of point signs through annotation also makes explicit some of the limitations of automated object detection paradigms, including the core assumptions of clear instance boundaries and minimal overlap, which are often unverified in practice. A YOLOv10-X detection model (Wang et al., 2024) is then trained on CaSiDD. Quantitative validation highlights high precision (81.1%), but lower recall (41.5%), which is primarily explained by interpretive divergences. For instance, repeated, short horizontal lines symbolizing wetlands (in the bottom example) might either be considered clusters of repeated, individual icons, or texture; likewise, marshes or hill icons are unclearly bounded, thus multiple object annotations and alternative prediction results could be considered valid. Overall, the high precision and lower recall advocates in favor of distributional research usages—such as comparing the relative prevalence of tree species in a certain region—over geographic reconstructive ones, as the model cannot reliably produce exhaustive feature layers.

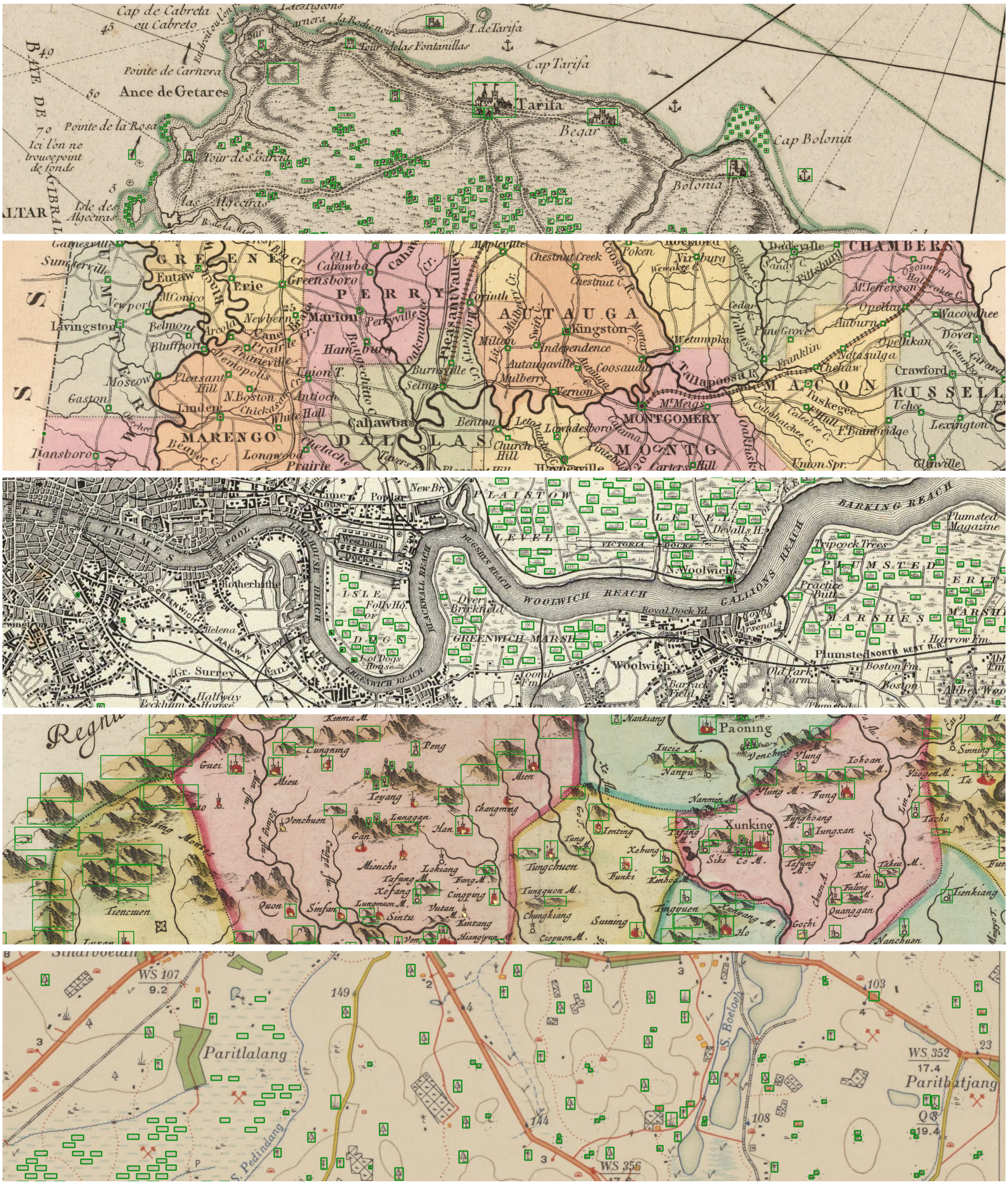

Figure 1: Example of sign detection on test set.

3 Encoding map signs

The second part of the paper investigates the methodological design space of visual sign encoding and validates it with respect to three dimensions: classification (semantics), attribution (form), and interpretability of the representation. The ambition is to achieve strong semantic and stylistic expressiveness, while explicitly suppressing material and collection-induced confounders. For each dimension, four distinct vision foundation models (VFMs) backbones are considered (Oquab et al., 2024; Ranzinger et al., 2026). Experiments assess the reaction of each VFM to a constrained domain-informed adaptation procedure. Two central challenges are the sensitivity of zero-shot representations to encroaching graphical elements in sign crops, which introduce noise into latent representations, and their contamination by highly predictive—yet uninformative—features, typically paper color or grain, background textures, and digitization artifacts. These weak, uninformative features act as confounders, artificially inflating classification accuracy or stylometric inference among maps sharing similar collection or production contexts while negatively impacting the interpretability of latent representations. These challenges are addressed by training a VFM adapter head to focus descriptions on morphological sign features while neutralizing background material cues. To produce the training targets, a lightweight foreground segmentation model is trained on synthetic samples. Then, the VFM adapter head learns to maximize cosine similarity between original sign patches embeddings, and preprocessed sign samples, using soft foreground predictions and color neutralization to focus feature representations. Three experiments are employed to validate VFM behavior under adaptation.

The first task assesses the ability to classify signs into 24 categories, such as tree, settlement, mill, vine, etc. The best precision and recall are achieved using an adapted RADIOv4 backbone ($P = 87.0\%$, $R = 87.4\%$).

In the second task, embeddings are used to compute map-to-map stylometric distances through a probabilistic sign density model, based on which clustering tendency is evaluated both within and across communities of mapmakers. Both adapted DINOv3 and base RADIOv4 achieve high map attribution rates, reaching 65.9% and 68.3%, respectively. Overall, the adapter often performs similar to base encoding for most backbones. Furthermore, the results indicate that local neighborhoods tend to be disrupted by adaptation. This suggests that local distances, plausibly relying on sensitive, non-structural graphical features, are negatively affected, as adaptation concentrates the representation on global visual components.

This interpretation is supported by the third experiment, which measures the linear association between embedding distances and explicit graphical features, describing color and sign shape (McConnell, 1986). Both preprocessing and adaptation appear to reduce the influence of background color and increase shape expressiveness. RADIOv4 exhibits maximal expressiveness for sign shape, especially with the adapter.

4 Exploring map signs

The automated detection and encoding of map signs open the way to the large-scale exploration and visualization of cartographic representation. Figure 2 presents a grid-constrained t-SNE (Maaten and Hinton, 2008) projection of 16,384 exemplars, extracted and sampled from 63 million signs across 99,715 historical maps. We briefly demonstrate how distributional analyses of this dataset can inform a computational history of cartography. For instance, by observing the frequency of sign occurrence, we identify moments of rupture, characterized by the rapid emergence of new signs and the disappearance of others. We also investigate the geographic structure of map semiotics across five countries, showing how signs tend to be locally differentiated, forming distinctive figurative cultures.

5 Conclusion

To conclude, this study introduces methods for scalable, generic evidence extraction from historical maps. It highlights the limits of object detection ontologies in semiotically ambiguous settings and proposes a controlled adaptation strategy for VFMs that generates more interpretable, less confounded embeddings, informing broader computational humanities practice beyond the specific case of cartography. The resulting framework supports distributional approaches to cartographic semiotics, interested in sign diffusion, technological disruptions and figurative shifts. Within geohistorical research, it could support comparative

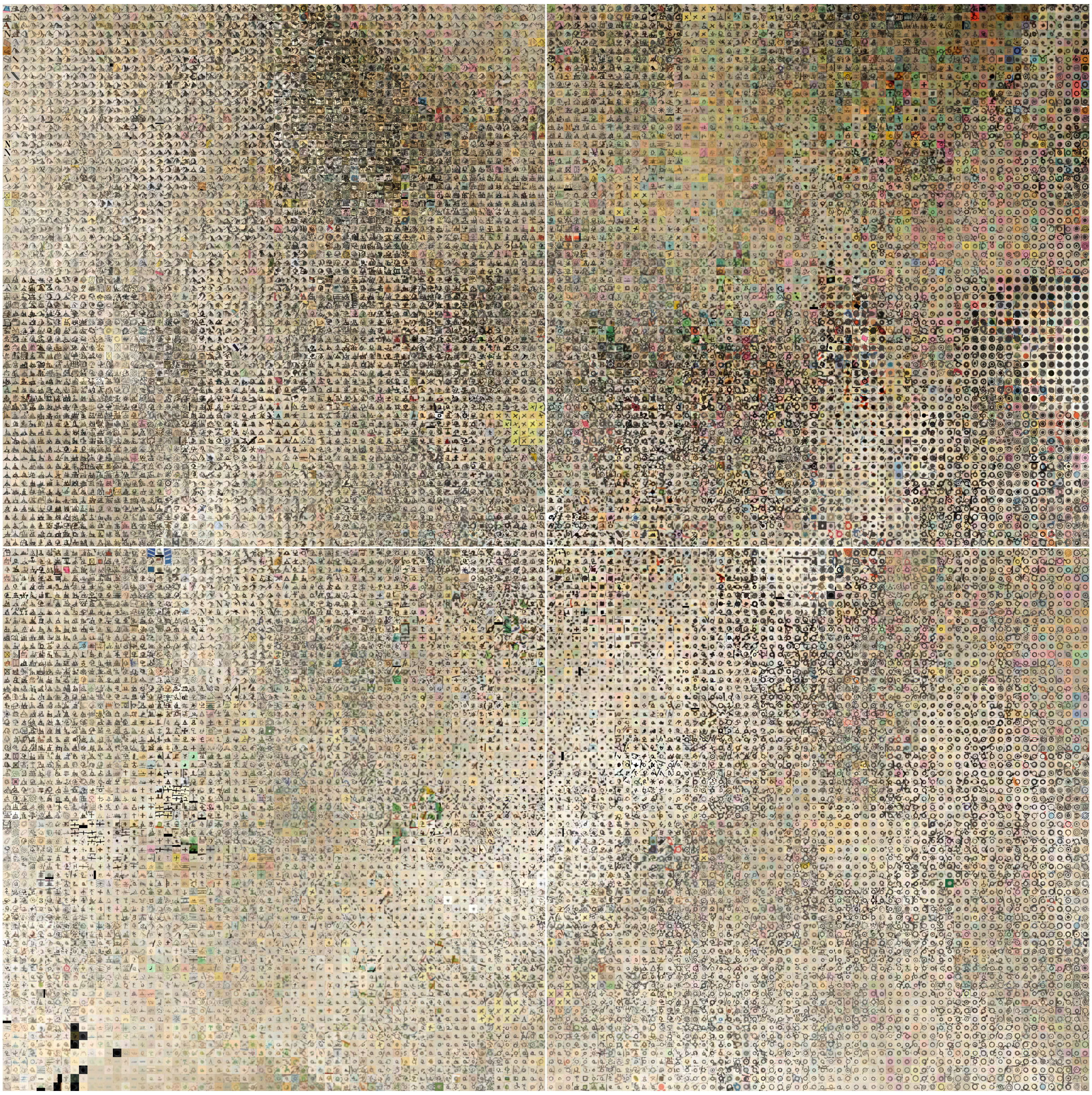

Figure 2: Visualization by grid-constrained t-SNE projection of 16,384 cartographic signs.

analyses of topics such as changes in vineyard extent following epiphytotic outbreaks, or enable density mapping of former peatlands with high carbon-storage capacity.

References

Laurens van der Maaten and Geoffrey Hinton. 2008. Visualizing Data using t-SNE. *Journal of Machine Learning Research* 9(11):2579–2605. http://www.jmlr.org/papers/v9/vandermaaten08a.html.

Robert K. McConnell. 1986. Method of and apparatus for pattern recognition. https://patents.google.com/patent/US4567610/en.

Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, Mahmoud Assran, Nicolas Ballas, Wojciech Galuba, Russell Howes, Po-Yao Huang, Shang-Wen Li, Ishan Misra, Michael Rabbat, Vasu Sharma, Gabriel Synnaeve, Hu Xu, Hervé Jegou, Julien Mairal, Patrick Labatut, Armand Joulin, and Piotr Bojanowski. 2024. DINOv2: Learning Robust Visual Features without Supervision. https://doi.org/10.48550/arXiv.2304.07193.

Remi Petitpierre and Jiaming Jiang. 2025. Cartographic Sign Detection Dataset (CaSiDD). https://doi.org/10.5281/zenodo.16278380.

Mike Ranzinger, Greg Heinrich, Collin McCarthy, Jan Kautz, Andrew Tao, Bryan Catanzaro, and Pavlo Molchanov. 2026. C-RADIOv4 (Tech Report). https://doi.org/10.48550/arXiv.2601.17237.

Eleanor S. Smith, Christopher Fleet, Stuart King, William Mackaness, Hannah Walker, and Catherine E. Scott. 2025. Estimating the density of urban trees in 1890s Leeds and Edinburgh using object detection on historical maps. *Computers, Environment and Urban Systems* 115:102219. https://doi.org/10.1016/j.compenvurbsys.2024.102219.

Charles Sutty and Bertrand Duménieu. 2024. CRH-EHESS/cocass-benchmarking. https://github.com/CRH-EHESS/cocass-benchmarking.

Ao Wang, Hui Chen, Lihao Liu, Kai Chen, Zijia Lin, Jungong Han, and Guiguang Ding. 2024. YOLOv10: Real-Time End-to-End Object Detection. https://doi.org/10.48550/arXiv.2405.14458.