The Journal of Open Source Software

MapReader: Open software for the visual analysis of maps

Rosie Wood [1], **Kasra Hosseini** [4], **Kalle Westerling** [1], **Andrew Smith** [1], **Kaspar Beelen** [1,3], **Daniel C. S. Wilson** [1], **and Katherine McDonough** [1,2¶]

1 The Alan Turing Institute, London, United Kingdom **2** Lancaster University, Lancaster, United Kingdom **3** School of Advanced Study, University of London, London, United Kingdom **4** Zalando SE, Berlin, Germany ¶ Corresponding author

DOI: 10.21105/joss.06434

Software

- Review
- Repository
- Archive

Editor: Elizabeth DuPre

Reviewers:

- @geekysquirrel
- @PipGrylls
- @jordibc

Submitted: 15 December 2023
Published: 26 September 2024

License
Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License (CC BY 4.0).

Summary

MapReader is an interdisciplinary software library for processing digitized maps (Hosseini et al., 2022a) and other types of images with two tasks: patch classification and text spotting. Patch classification works by 'patching' images into small, custom-sized cells which are then classified according to the user's needs. Text spotting detects and recognizes text. MapReader offers a flexible pipeline which can be used both for manual annotation of small datasets as well as for computer-vision-based inference of large collections. As an example, in Hosseini et al. (2022a), we annotated 62,020 patches, trained a suite of computer vision models and performed model inference on approximately 30.5 million patches.

MapReader's approach was inspired by methods in biomedical imaging, which were adapted for use by historians, and it is suitable for a wide range of applications in image analysis: it has, for example, been applied to an image classification problem in plant phenotype research (Corcoran et al., n.d.). This cross-pollination between the humanities and the natural sciences was made possible by the open and reproducible research methods at the heart of MapReader.

MapReader pioneers a methodological shift in how historians interact with maps as primary sources. Sustained engagement with big collections of maps rarely moves beyond analysis of cartographic history. To change this, MapReader encourages historians to reflect on the content of maps and is designed to facilitate linking datasets representing visual map content with other historical geospatial data to enable spatial historical research.

In this paper, we present the MapReader release at the conclusion of the Living with Machines project, which supported the initial development of the software and associated historical research. This release represents the culmination of extensive work to improve MapReader's usability among historians, especially through clear documentation and tutorials.

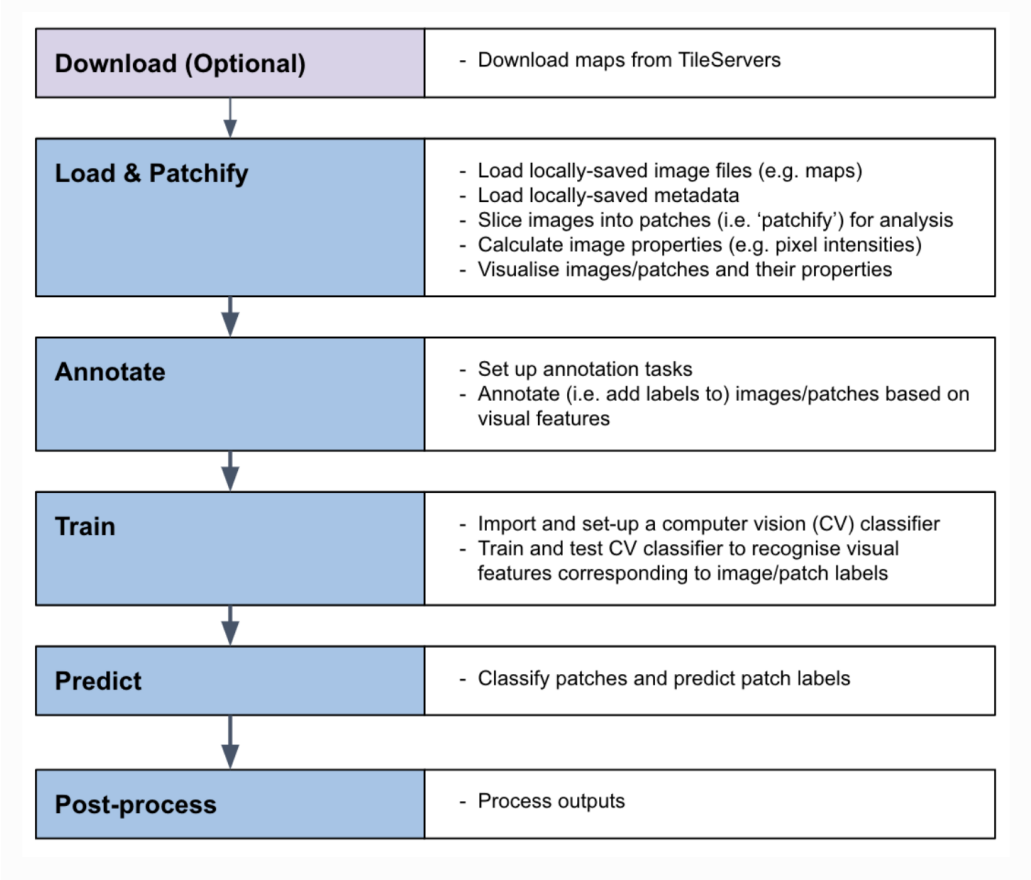

Figure 1: MapReader modules and input-outputs. Credit: Rosie Wood.

Statement of need

Since the 1990s, map libraries have been scanning maps and creating digital collections of these images (Hosseini et al., 2021). As of 2023, there are more than a million images of maps in digital libraries and archives around the world, and yet it is very difficult for anyone to do more than browse them in a web viewer.

MapReader makes it possible to ask questions of thousands of digitized maps at a time, a fundamentally different intellectual experience from both the traditional manner of viewing a few maps at a time on a reading room table as well as the act of visually scanning digital files sequentially. We used MapReader to process a collection of approximately 16,000 nineteenth-century Ordnance Survey map sheets (approximately 30.5 million patches) covering England, Wales and Scotland (Hosseini et al., 2022a). Inspired by the possibility of seeing a series of maps stitched together in seamless layers—such as in the National Library of Scotland's Ordnance Survey map viewing interface—MapReader takes the next step by transforming the experience of working with maps from surface exploration to critical investigation (Hosseini et al., 2021).

Related Work

MapReader is among the first end-to-end pipelines for processing historical maps and other images. It has been designed to lower barriers to experimenting with computer vision and answering research questions about large image datasets. Other projects are emerging which

are performing similar research tasks with the visual content in historical map collections (Combes et al., 2022; Petitpierre, 2021), and of course other tools, like the Distant Viewing Toolkit (Arnold & Tilton, 2020), address similar needs for other kinds of media.

In addition, as part of a collaboration between Machines Reading Maps and the David Rumsey Historical Map Collection, the Knowledge Computing Lab released mapKurator (Li et al., 2020) - the first text spotting pipeline for maps - which takes map image input and and returns polygons and text transcriptions in geojson format. Now, MapReader also incorporates this text spotting task in addition to its original patch classification task.

Documentation

MapReader aims to build computational skills among historians. Our extensive work on documentation and training, including substantial updates to MapReader since Hosseini et al. (2022a), reflect this commitment. As historians explore the possibilities of computational methods for novel historical research, MapReader models how computational tools can unlock difficult-to-use primary sources and how we can embrace open research practices as a way to encourage learning. We welcome contributions and requests for new documentation or tutorials.

Our documentation aims to provide all information needed to set up and run MapReader. This includes:

- A basic introduction to the software and its origins;
- A project Curriculum Vitae - i.e. papers, talks, workshops, etc. delivered by the MapReader team;
- Information about MapReader events and activities, where the community can engage with MapReader;
- Guidance for users who are new to coding;
- Guidance on input - e.g. what kind of maps and which formats work well in MapReader as well as how to set up files;
- A step-by-step guide to installing and running MapReader;
- Worked examples - jupyter notebooks demonstrating MapReader use cases with data provided, e.g. Hosseini et al. (2022b);
- An API Reference;
- A Code of Conduct and Inclusivity statement; and
- Guidance for contributing to MapReader.

Conclusion

Through its conceptual approach, modular structure, documentation, and worked examples, MapReader enables researchers to ask questions of large collections of maps. It represents a novel approach to digitizing map content, one which intentionally prevents the collection of overly precise data from cartographic documents. MapReader embraces a humanistic approach to data creation and curation, offering an alternative or complement to pixel-level image segmentation.

Acknowledgements

This work was supported by Data/Culture (AHRC grant AH/Y00745X/1), Living with Machines (AHRC grant AH/S01179X/1), and The Alan Turing Institute (EPSRC grant EP/N510129/1). Living with Machines, funded by the UK Research and Innovation (UKRI) Strategic Priority Fund, was a multidisciplinary collaboration delivered by the Arts and Humanities Research Council (AHRC), with The Alan Turing Institute, the British Library and the Universities

of Cambridge, East Anglia, Exeter, and Queen Mary University of London. Maps and their metadata in MapReader are reproduced with the permission of the National Library of Scotland (https://maps.nls.uk/index.html). We also wish to thank participants in events in 2023 and 2024 who provided feedback on using MapReader.

Contribution Statement

Katherine McDonough wrote and revised this article, with substantial contributions from Daniel C.S. Wilson and Rosie Wood. Andy Smith, Kalle Westerling, Kaspar Beelen and Kasra Hosseini reviewed the final manuscript. Please see contributions to the MapReader software library at https://github.com/maps-as-data/MapReader#contributors, including work from all named authors.

References

Arnold, T., & Tilton, L. (2020). Distant Viewing Toolkit: A Python Package for the Analysis of Visual Culture. *Journal of Open Source Software*, *5*(45), 1800. https://doi.org/10.21105/joss.01800

Combes, P.-P., Gobillon, L., & Zylberberg, Y. (2022). Urban economics in a historical perspective: Recovering data with machine learning. *Regional Science and Urban Economics*, *94*, 103711. https://doi.org/10.1016/j.regsciurbeco.2021.103711

Corcoran, E., Hosseini, K., Siles, L., Kurup, S., & Ahnert, S. (n.d.). *Dynamic phenotyping of whole oilseed rape (brassica napus) plants from images collected under controlled conditions*.

Hosseini, K., McDonough, K., Van Strien, D., Vane, O., & Wilson, D. C. S. (2021). Maps of a Nation? The Digitized Ordnance Survey for New Historical Research. *Journal of Victorian Culture*, *26*(2), 284–299. https://doi.org/10.1093/jvcult/vcab009

Hosseini, K., Wilson, D. C. S., Beelen, K., & McDonough, K. (2022a). MapReader: A computer vision pipeline for the semantic exploration of maps at scale. *Proceedings of the 6th ACM SIGSPATIAL International Workshop on Geospatial Humanities*, 8–19. https://doi.org/10.1145/3557919.3565812

Hosseini, K., Wilson, D. C. S., Beelen, K., & McDonough, K. (2022b). *MapReader_data_SIGSPATIAL_2022*. Zenodo. https://doi.org/10.5281/ZENODO.7147906

Li, Z., Chiang, Y.-Y., Tavakkol, S., Shbita, B., Uhl, J. H., Leyk, S., & Knoblock, C. A. (2020). An automatic approach for generating rich, linked geo-metadata from historical map images. *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 3290–3298. https://doi.org/10.1145/3394486.3403381

Petitpierre, R. (2021). *Neural networks for semantic segmentation of historical city maps: Cross-cultural performance and the impact of figurative diversity*. https://doi.org/10.48550/ARXIV.2101.12478