MapReader: Open software for the visual analysis of maps

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

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

Summary

MapReader is an interdisciplinary software library for processing digitized maps (Hosseini et al., 2022a), but also other types of images, by 'patching' them into small, custom-sized cells, which are then classified according to the user's needs. MapReader thus 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 utilized MapReader's interface to manually annotate 62,020 patches, used its functionalities to train 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., 2023). 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.

In this paper, we present the MapReader release at the conclusion of the Living with Machines project, which supported the development of the software and associated historical research. This release represents the culmination of extensive work to improve MapReader's usability, especially through clear documentation and tutorials.
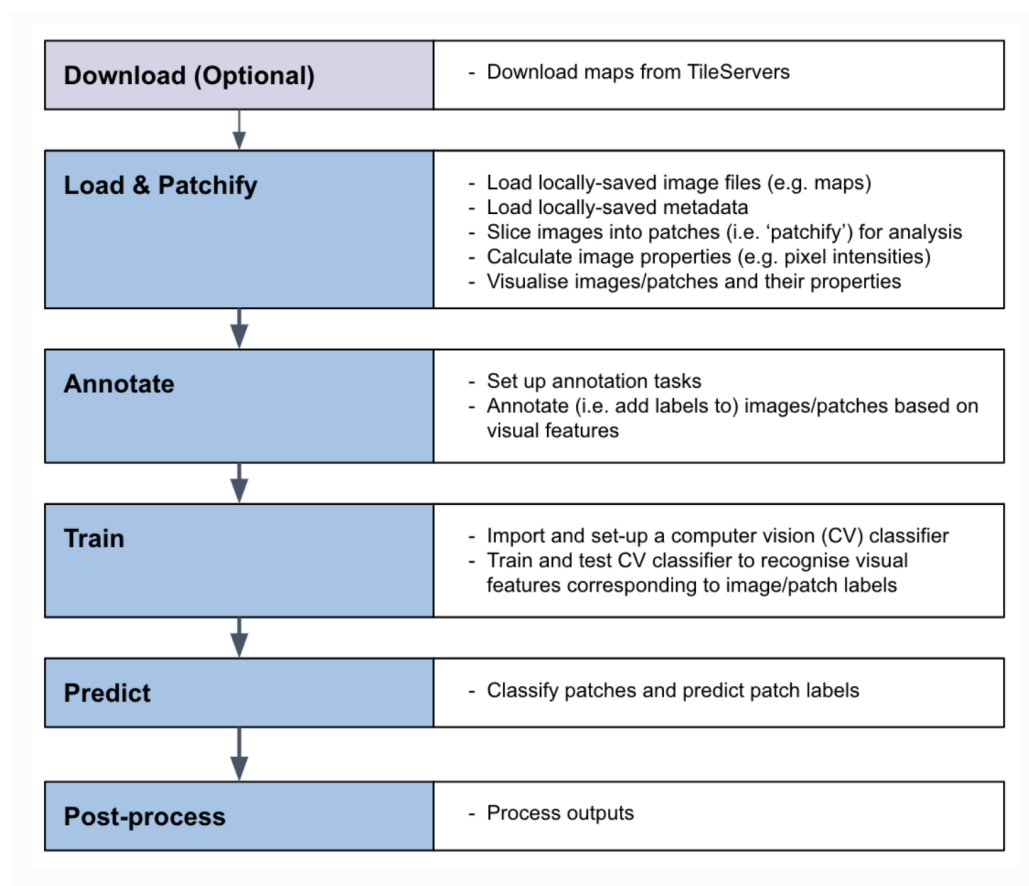
Figure 1: MapReader modules and input-outputs. Credit: Rosie Wood.

Statement of need

Since the 1990s, map libraries have been scanning maps and creating digital collections of these images (Hosseini et al., 2021). In 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 and the act of visually scanning digital files sequentially. As an example, we used MapReader to process a collection of ~16K nineteenth-century Ordnance Survey map sheets (~30.5M patches) covering England, Wales and Scotland (Hosseini et al., 2022a). Inspired by the possibility of seeing series maps stitched together in seamless layers, such as the National Library of Scotland 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 pipeline for processing historical maps and other images that was designed to lower barriers to experimenting with computer vision in 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. As part of a collaboration between Machines Reading Maps and the David Rumsey Historical Map Collection, the Knowledge Computing Lab released mapKurator (Lab, n.d.), which takes map image input and creates geojson files of all text on each map.

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 includes: - About MapReader: A basic introduction to the software and its origins - Events: Activities where the community can engage with MapReader - Project Curriculum Vitae: Papers, talks, workshops, etc. delivered by the MapReader team - Coding Basics: Skills and tools that are useful for getting started with MapReader - Installation Instructions: How to install MapReader - Input Guidance: What kind of maps and which formats work well in MapReader - User Guide: Walkthrough of how to run MapReader - Worked Examples: Jupyter notebooks demonstrating uses of MapReader for specific cases (with data, e.g. Hosseini et al. (2022b)) - API Reference: Auto-generated API reference documentation - Code of Conduct and Inclusivity: Our approach to ethical and inclusive conduct - Contribution Guide: How to engage with MapReader - Developer's Guide: How to update the MapReader version number and upload to package managers

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

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/Living-with-machines/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. (2023). Automated extraction of dynamic phenotype data from whole plant images collected under controlled conditions. *Frontiers in Plant Science*, In preparation.

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

Lab, K. C. (n.d.). mapKurator. In *Github repository*. Github. Retrieved November 23, 2023, from https://knowledge-computing.github.io/mapkurator-doc/#/

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