MapReader

Katie McDonough, Daniel Wilson, Rosie Wood

Environmental Digital Humanities Workshop

29 January 2025
University of Manchester

Our Partners

The Alan Turing Institute
British Library
University of Cambridge
University of East Anglia
University of Exeter
Queen Mary University of London

Our Funders

Arts & Humanities Research Council
UK Research and Innovation

Workshop material and research applications co-developed with:

Rosie Wood, Turing

Kalle Westerling, Turing

Daniel Wilson, Turing

Kaspar Beelen, SAS, University of London

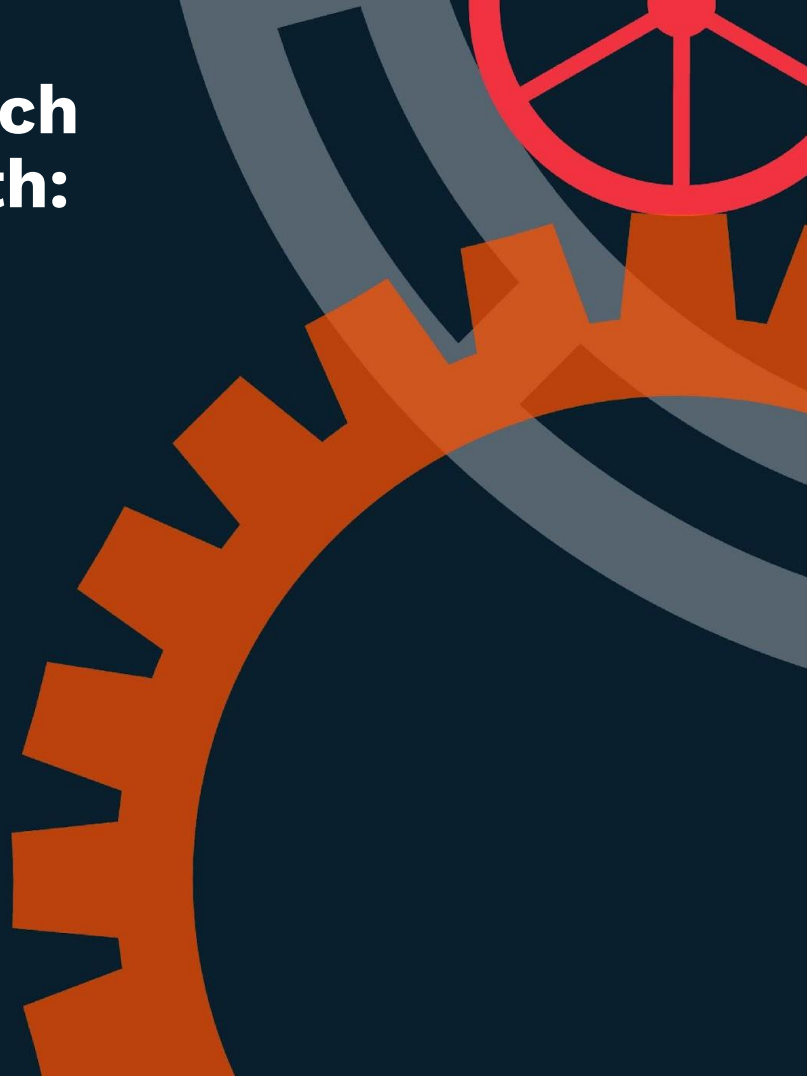
Tim Hobson, Turing

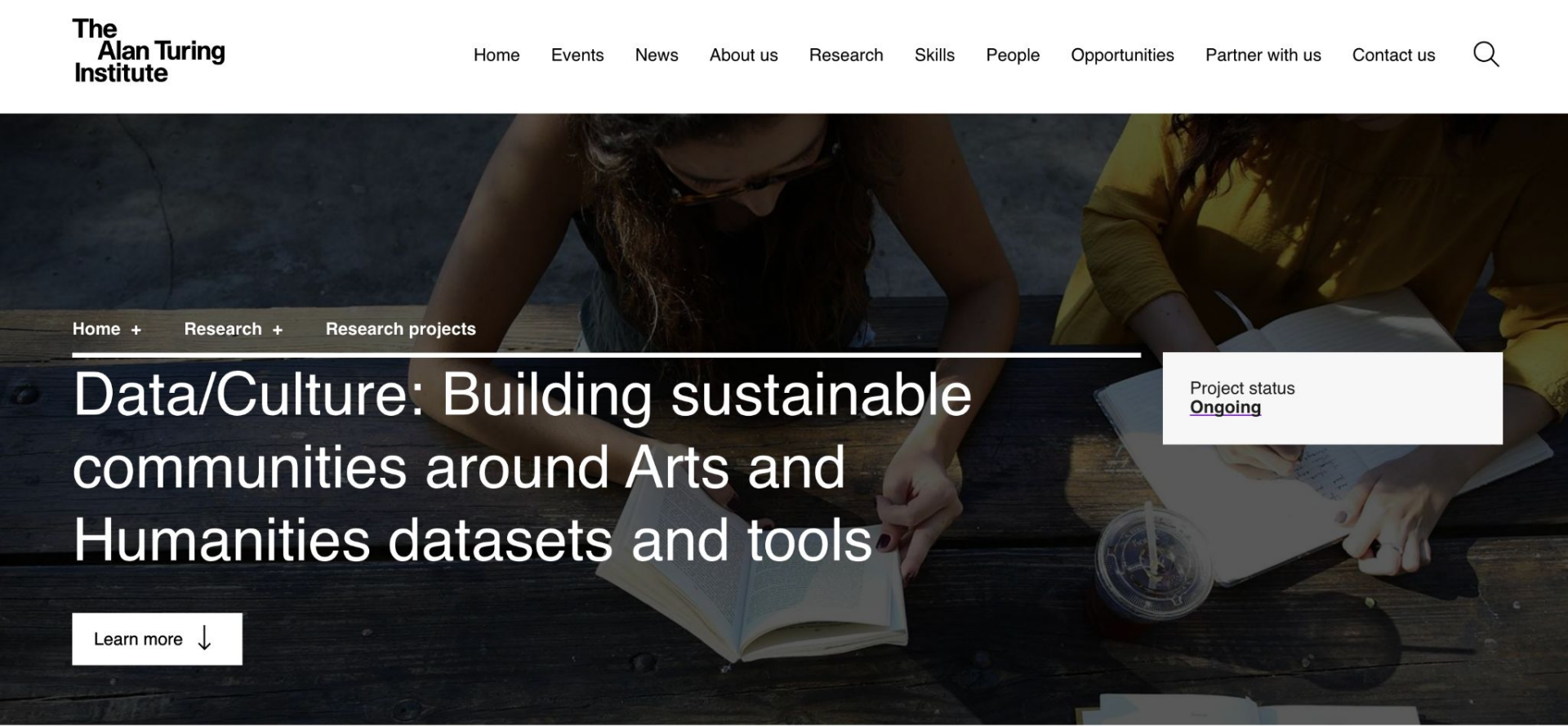
Kasra Hosseini, Zalando (formerly Turing)

Josh Rhodes, Durham

Jon Lawrence, Exeter

Andy Smith, Turing





https://www.turing.ac.uk/research/research-projects/dataculture-building-sustainable-communities-around-arts-and-humanities

Why MapReader?

1888

How do researchers work with maps?

Digitized maps can be more than sheets to browse in a virtual reading room. *But how?*

Ordnance Survey maps of England, Wales, and Scotland

6 inches to 1 mile
1888-1913
(2nd edition)

~15K sheets

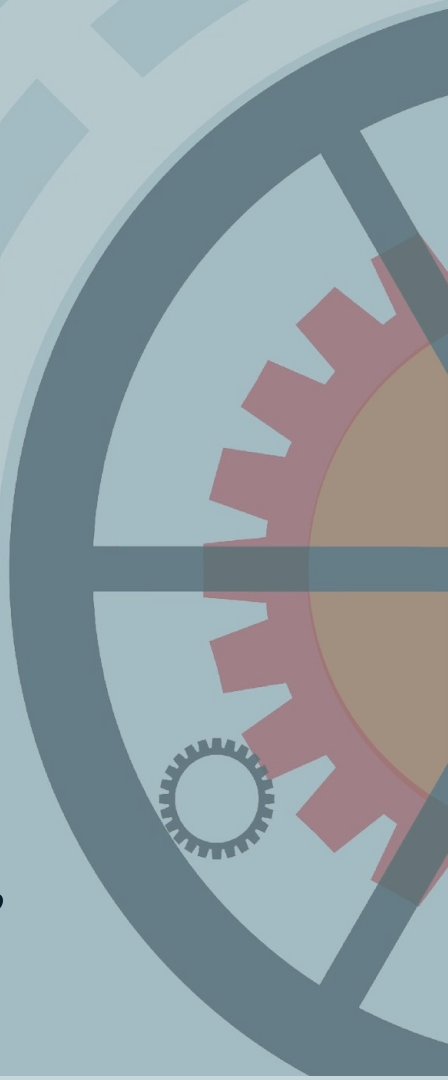
Credit: Olivia Vane for visualisation, National Library of Scotland for map images & metadata

Make trustworthy claims based on *thousands* of maps

case studies → ‘high resolution’ local archival research or anecdotes from printed materials

aggregated statistics → ‘low resolution’ regional/national

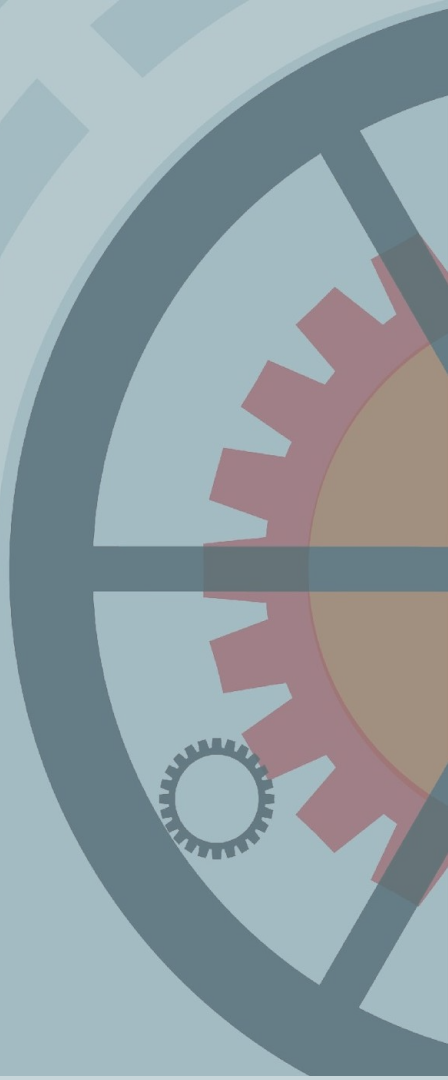
Ordnance Survey (series) maps as non-aggregated, high-resolution, national-coverage sources



You Cannot Ground Truth the Past

Historical OS maps tell us how Victorians represented the British landscape, and how that landscape was changing, but they are not a ‘ground truth’.

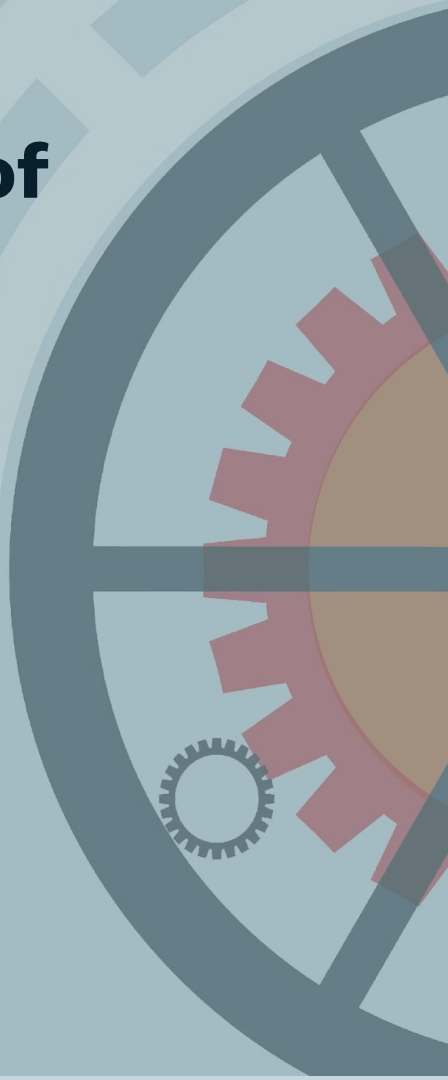
We want to ***use CV to advance interpretation, not define truth.***



Step away from the GIS paradigm of vectorizing everything on a map

Stop historical **data “mining”/“extraction”** and insisting on overly-precise data

Start developing methods that permit **critical distance between scholars and sources** & question-appropriate data collection and exploration



Solution: ‘Patches’ as a new shape for historical research

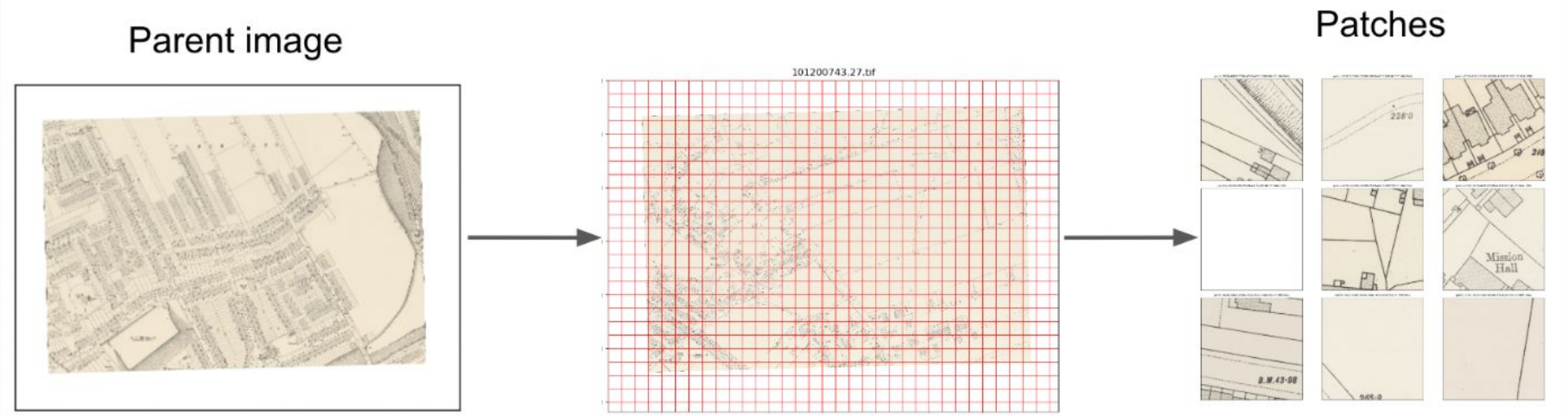


Image Classification for Map Patches

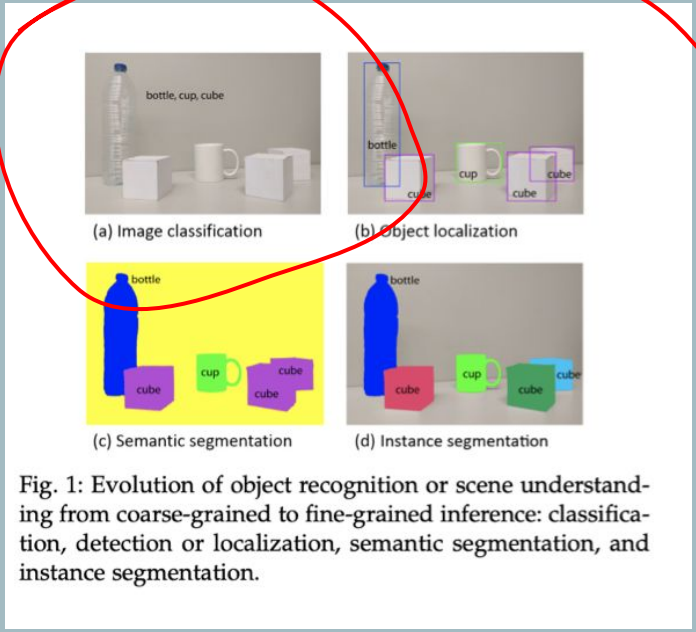
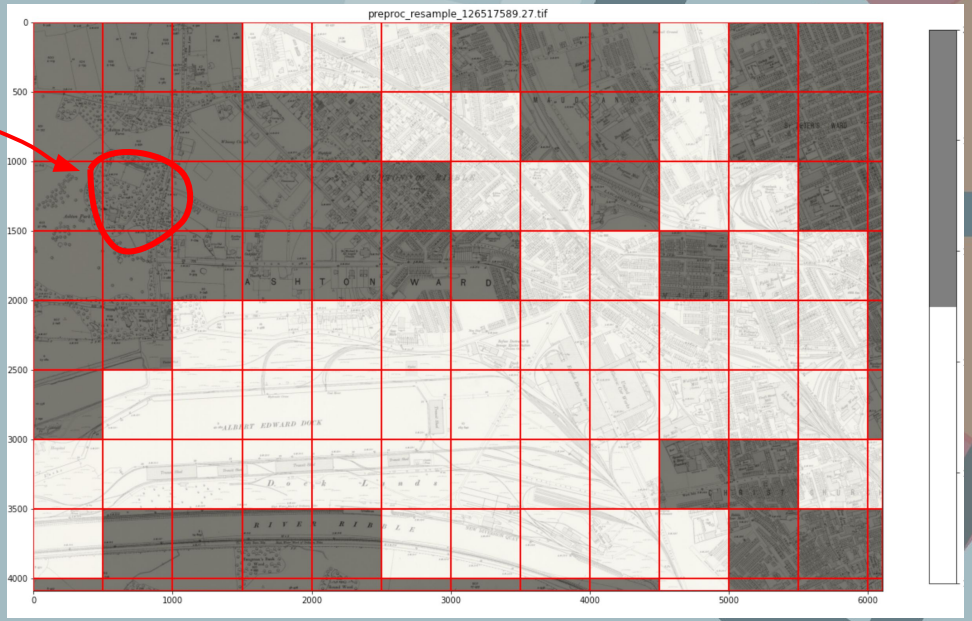


Fig. 1: Evolution of object recognition or scene understanding from coarse-grained to fine-grained inference: classification, detection or localization, semantic segmentation, and instance segmentation.

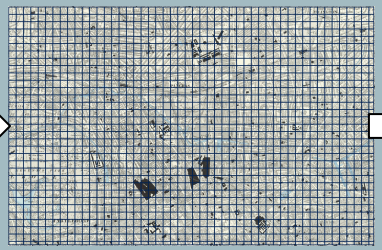
https://arxiv.org/pdf/1704.06857.pdf



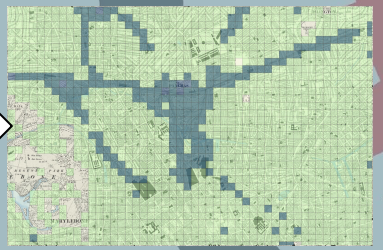
Map patches annotated as training data

Classification pipeline

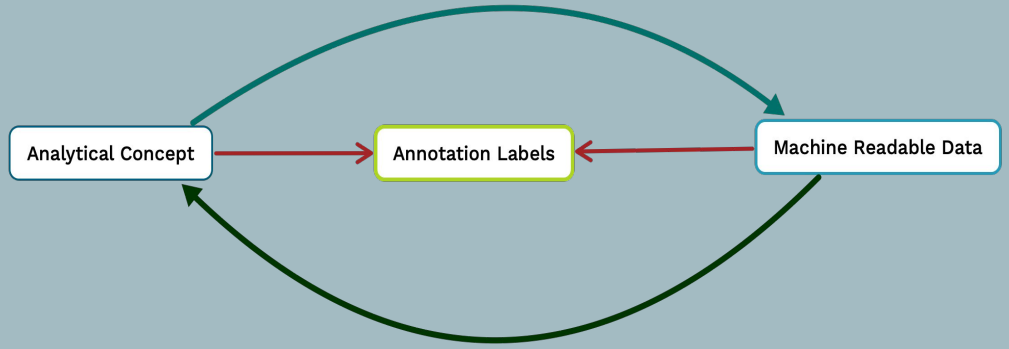
Download	Load	Annotate	Classify
Optional: - Download maps	- Load image files - Load metadata - Patchify	- Annotate patches based on visual features	- Train classifier - Predict labels

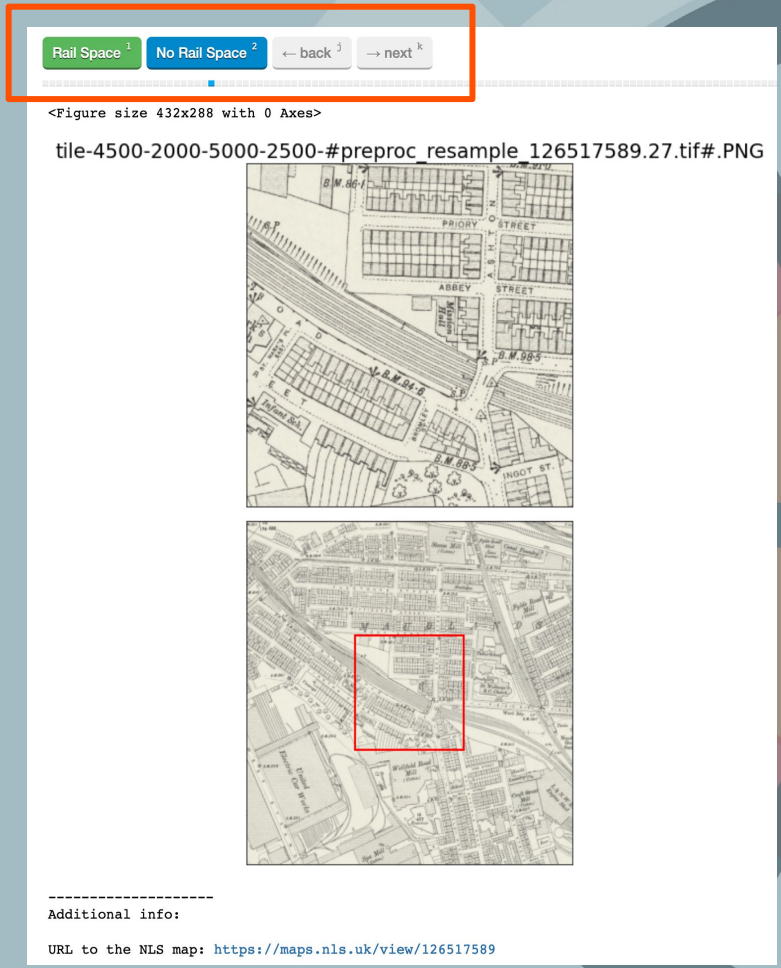


railspace



Annotating patches: What is a good label?





What was it like to live *with* railways?

Railspace as the total footprint of rail infrastructure

- Railspace label inferred on ~15K late nineteenth-century 6” OS maps (NLS)
- ~30.5 million 100x100m patches
- ~62k expert-annotated patch dataset now available on Zenodo/Hugging Face

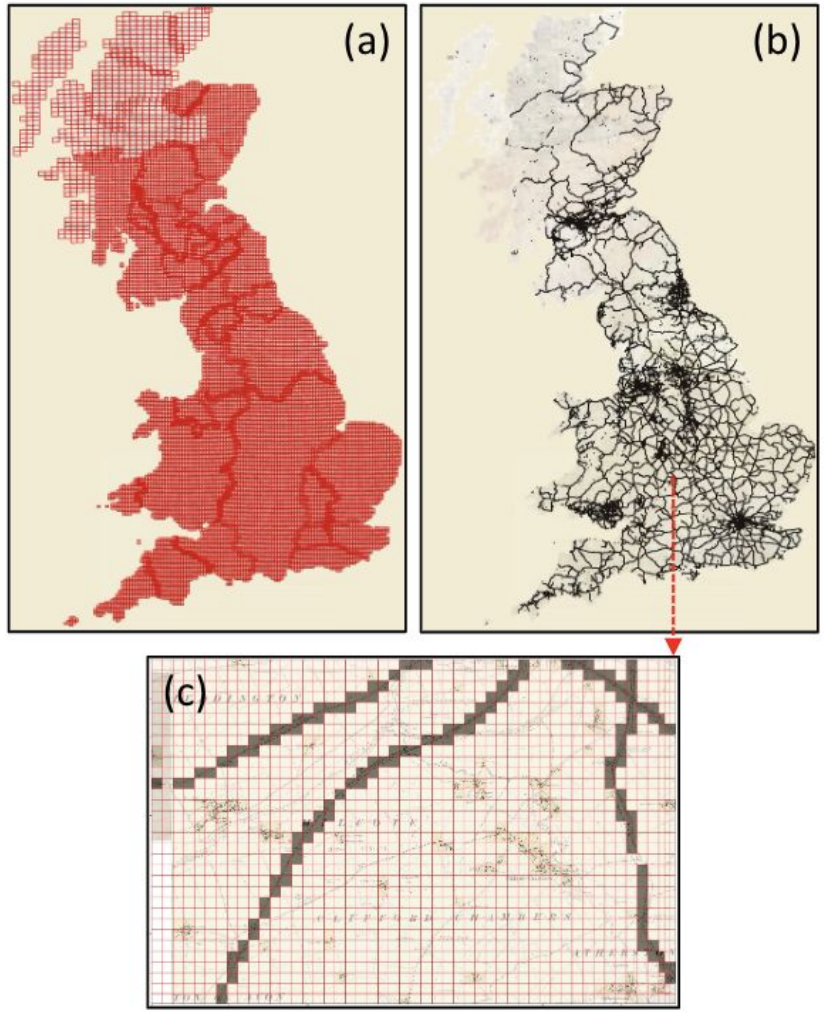
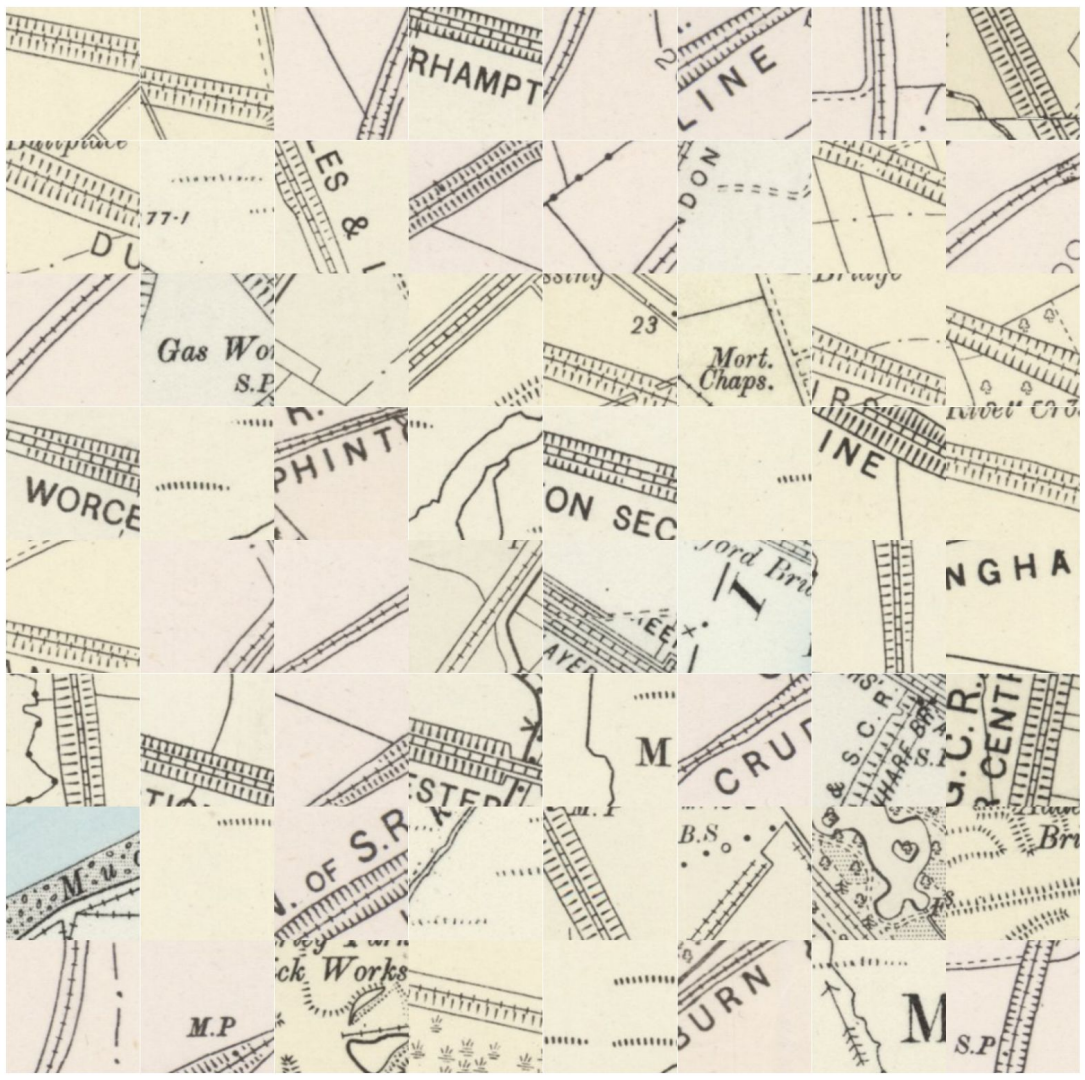


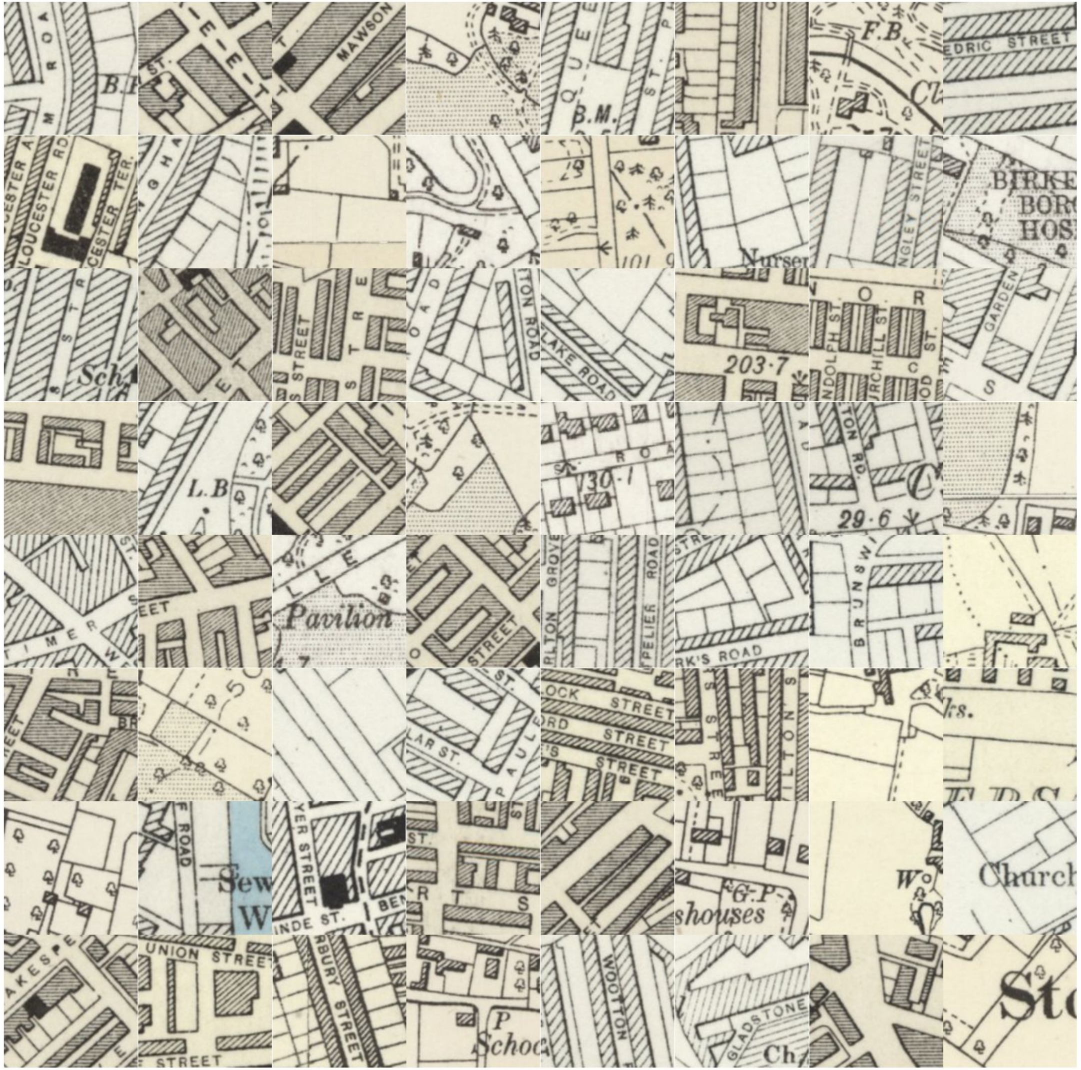
Image credit: Kasra Hosseini

Railspace (rural)



Railspace (urban)

Buildings



MAWSON
EDRIC STREET
B.M.
F.B.
BIRKE
BORO
HOS
Nurser
LOUCESTER RD
LAKE ROAD
203·7
30·1
29·6
L.B.
Pavilion
GROVE
RK'S ROAD
BRUNSW
LOCK STREET
ORD STREET
LAR ST.
Sew
W
INDE ST.
G.P.
shouses
Church
UNION STREET
WOOTTON
GLADSTONE
Schoo

Living with Machines

Computational Histories of the Age of Industry

by Ruth Ahnert, Emma Griffin, Jon Lawrence, The Living with Machines Team

ABOUT DIGITAL EDITIONS

Living With Machines is a data-driven history of the coming of the machine age in Britain in the long nineteenth century. Featuring an innovative open access edition enhanced with interactive maps, datasets and visualisations, digital notebooks, video, audio and images, this book harnesses the combined power of massive digitised historical collections and computational analytical tools to examine the ways in which technology altered the very fabric of human existence on a hitherto unprecedented scale.

Beyond the Tracks: re-connecting people, places and stations in the history of late-Victorian railways

Joshua Rhodes, Jon Lawrence, Kaspar Beelen, Katherine McDonough, Daniel C.S. Wilson

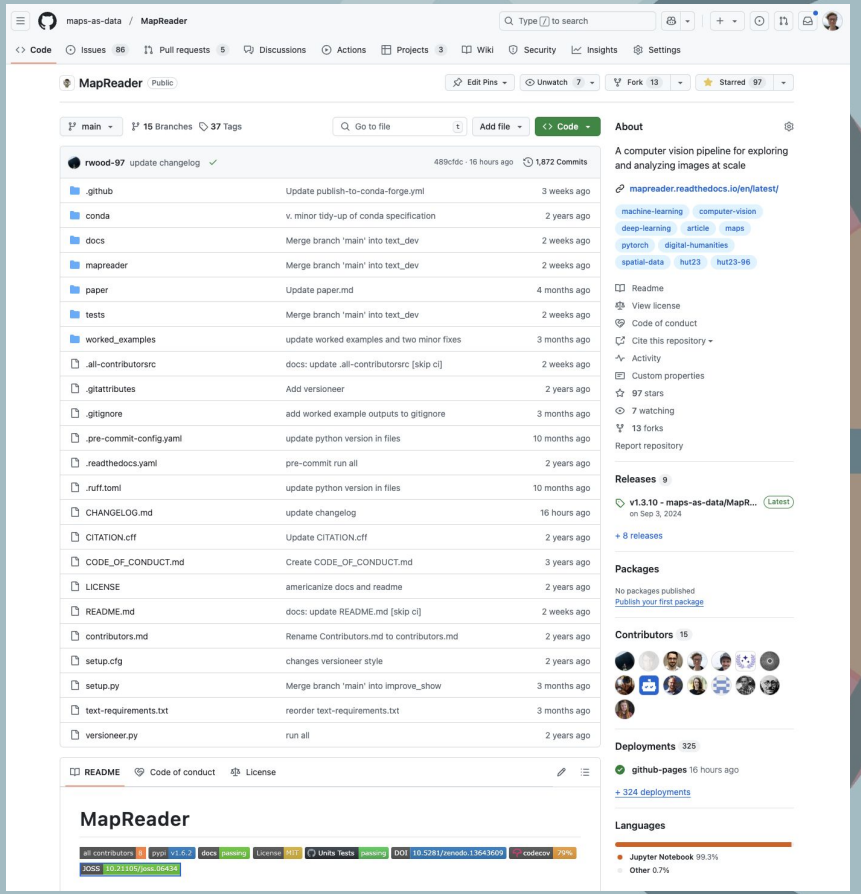
MapReader in 2025

MapReader performs 2 tasks:

1. Classify map regions (patches) with concepts that have a **visual** signal

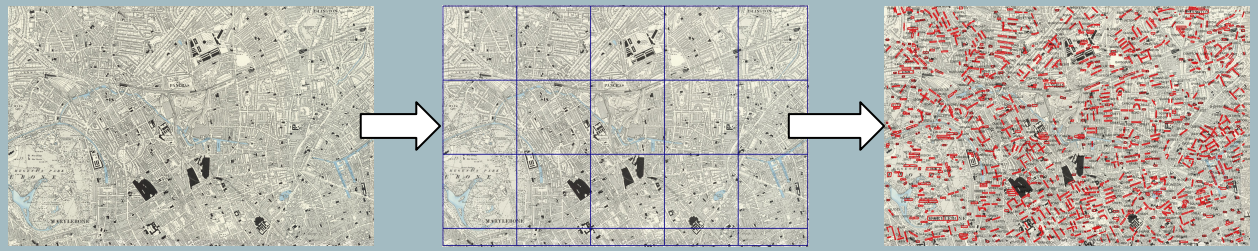
2. Find and transcribe **text**

github.com/maps-as-data/MapReader



Text spotting pipeline

Download	Load	Spot text
Optional: - Download maps	- Load image files - Load metadata - Patchify	- Detect and recognise text



MapReader & AI

MapReader is a sandpit for thinking with map content and metadata to critically create structured data for onward linking and analysis.

We have privileged software design and methods that:

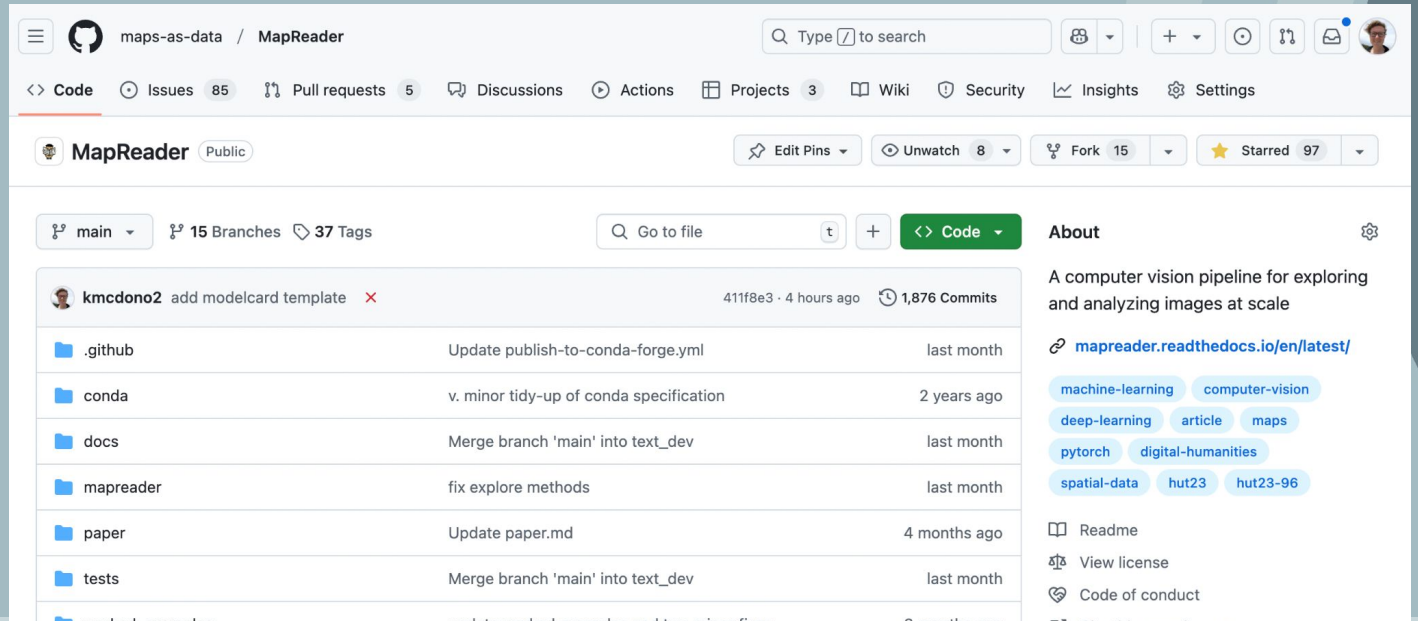
- **Illuminate understudied cartographic elements, like text**
- **Encourage interpretation over extraction, with patches**
- **Foster stronger connections with collections & curators**

https://github.com/maps-as-data/MapReader

Using MapReader

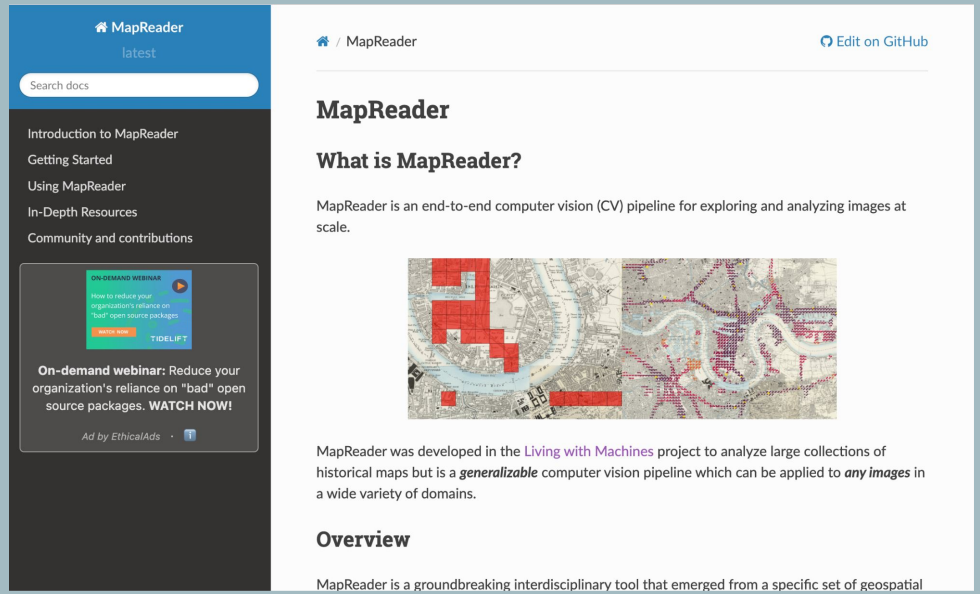
MapReader Github Repository

https://github.com/maps-as-data/MapReader

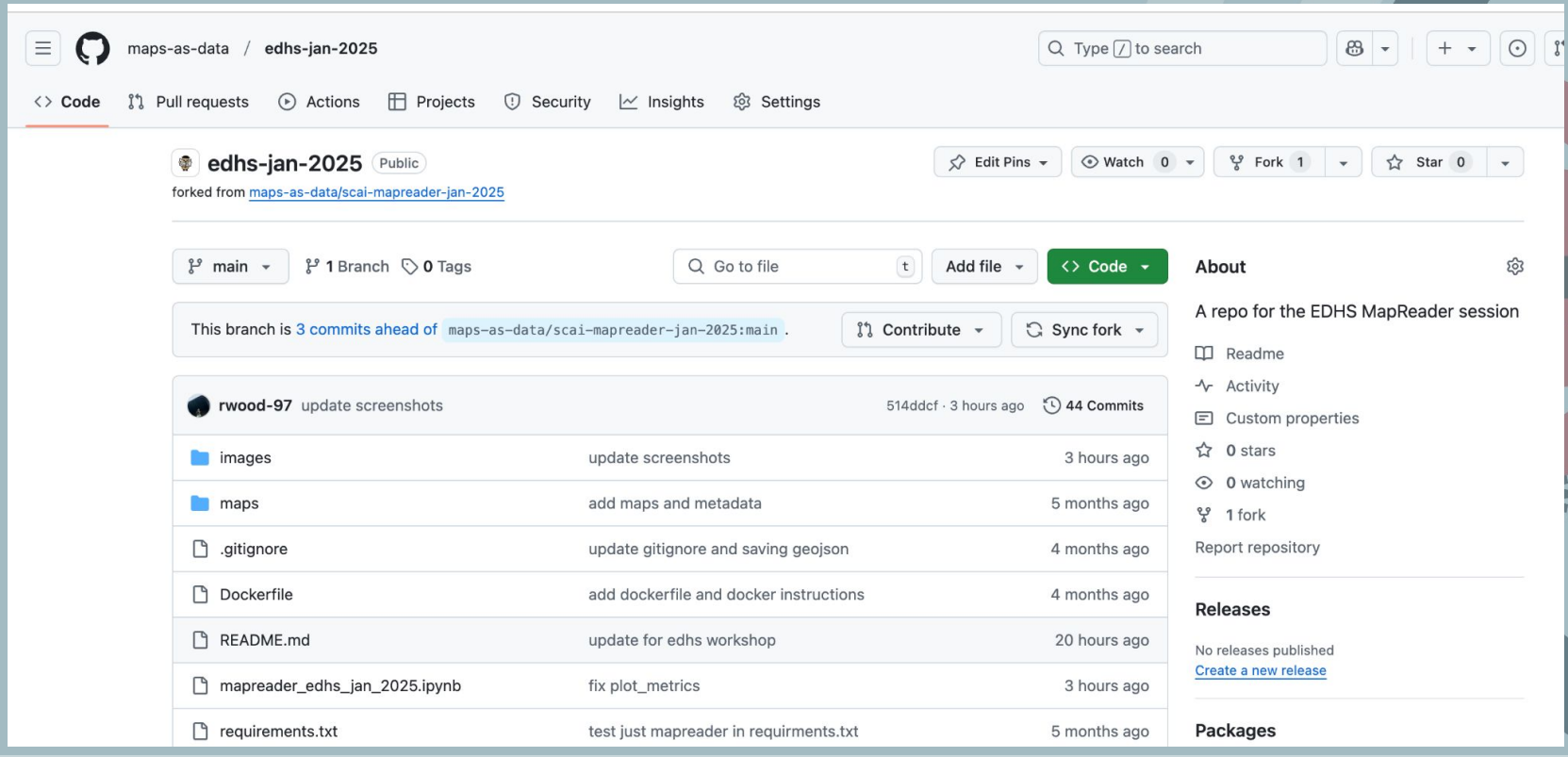


MapReader Documentation

https://mapreader.readthedocs.io/en/latest/

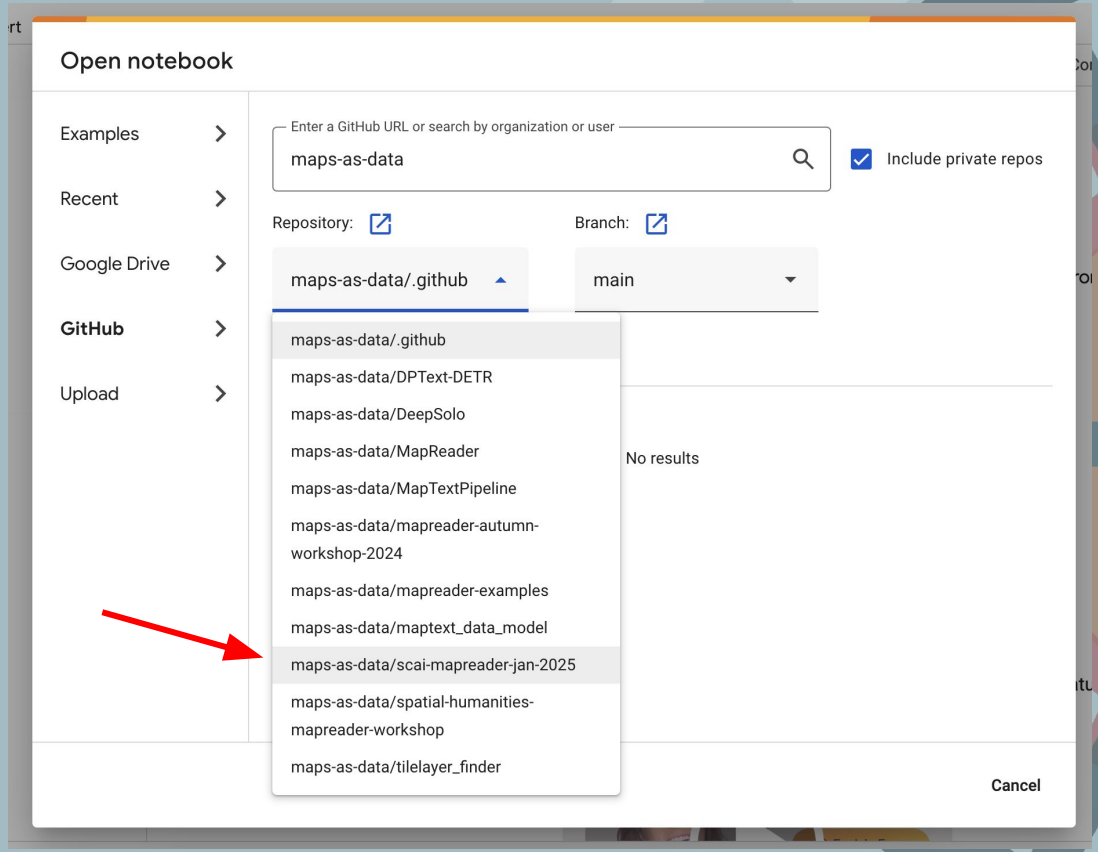


https://github.com/maps-as-data/edhs-jan-2025



Setting Up

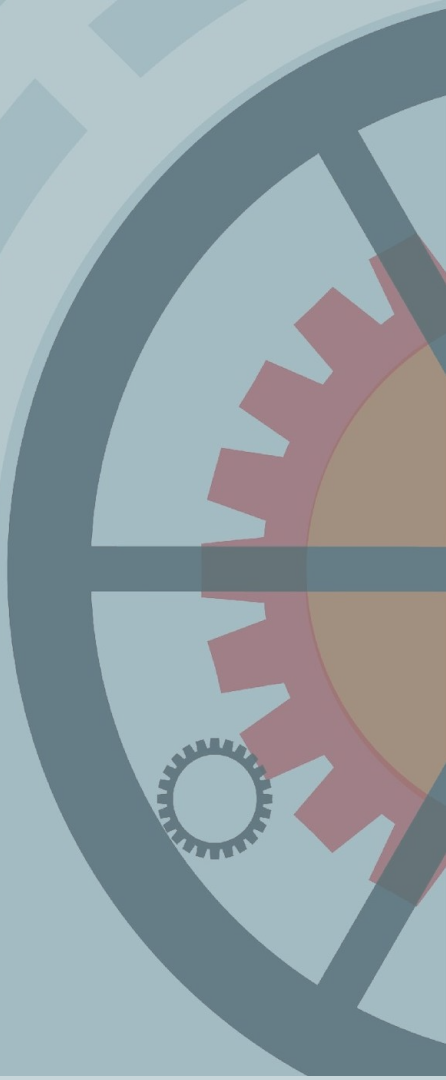
Follow the instructions for Google Colab in the README



Jupyter Notebook Basics

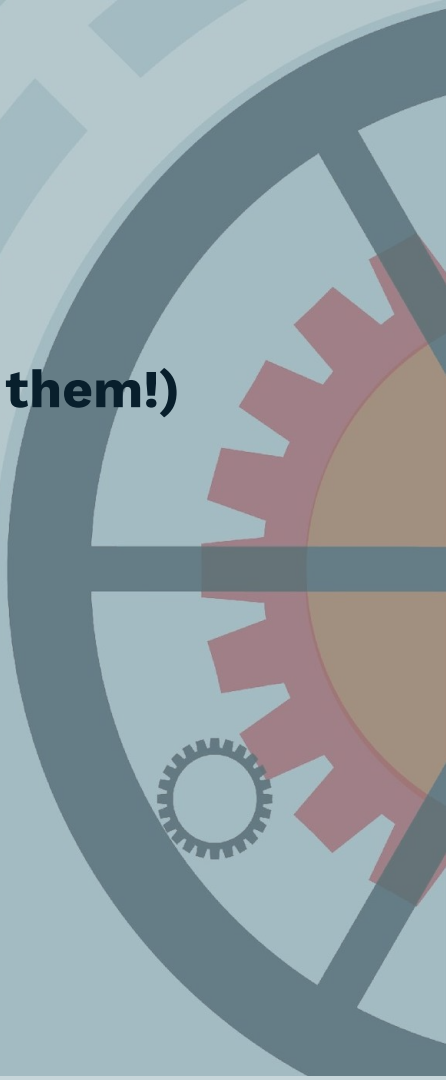
- **Two different cell types - Code or Text**
- **Running cells - Press "▶" or "⌘/Ctrl/Shift+Enter"**

WARNING: Google Colab will time out after a while so make sure you save your work!

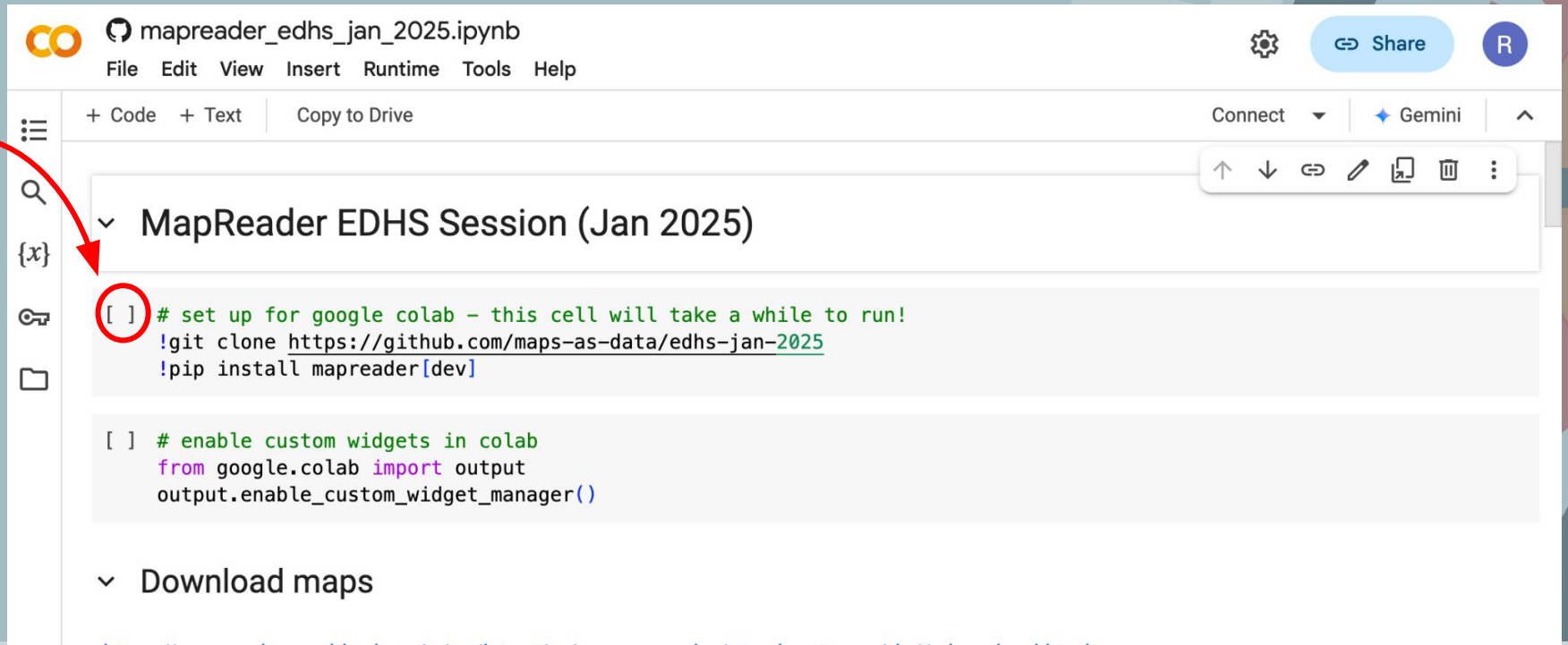


Notebook Sections

- (No need to download maps, we have provided them!)
- Install dependencies
- Load & Patchify maps
- Annotate patches with your label of choice
- Fine-tune a model
- Review model performance
- Infer labels on remaining patches
- Export labeled patch data



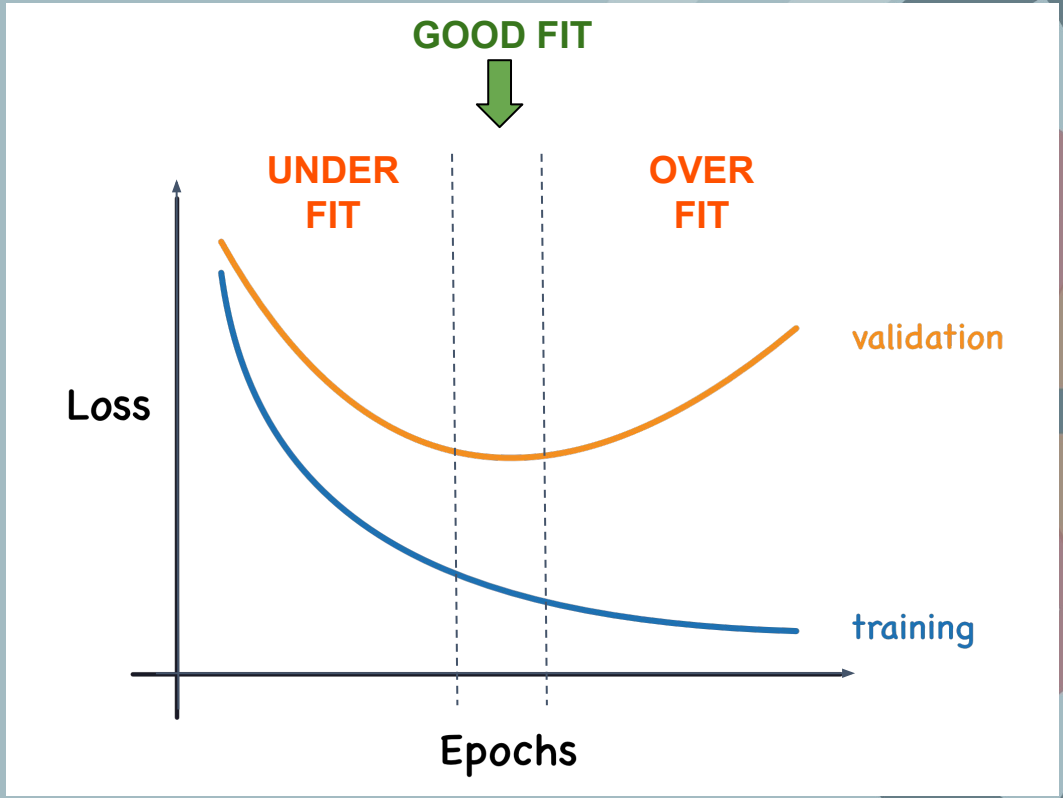
Open notebook & run first code cell



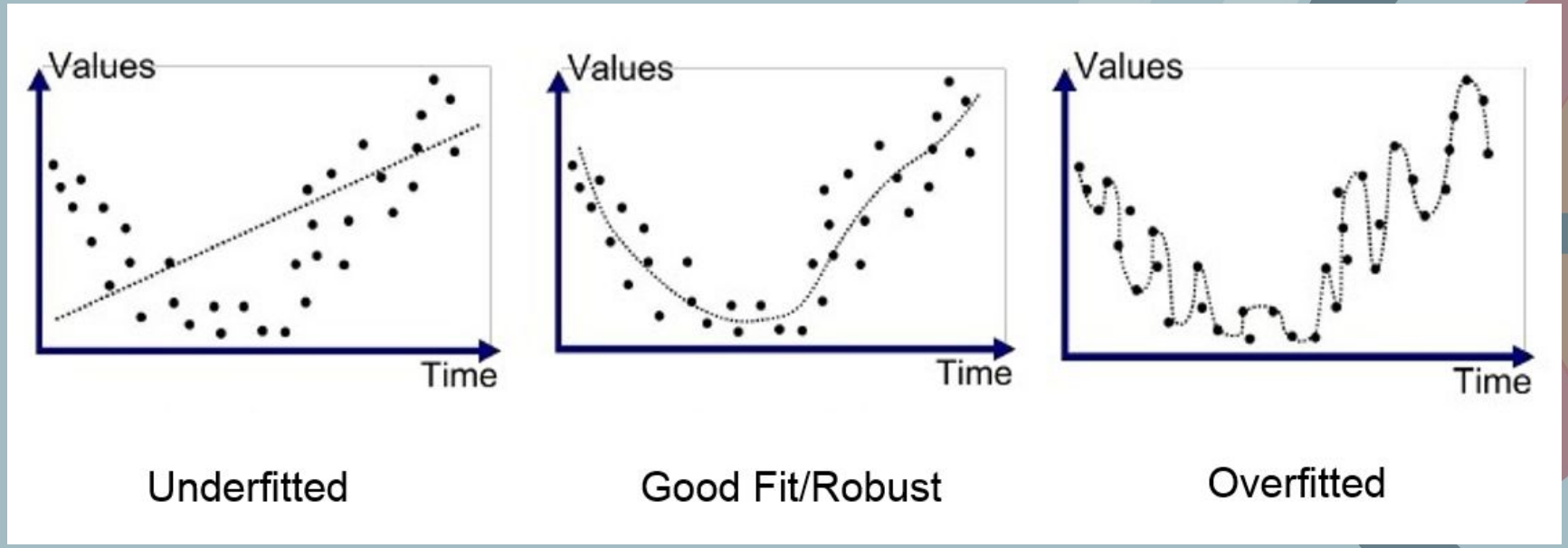
Training

Annotations split into train, validation and test datasets

Goal is to minimise both training and validation losses



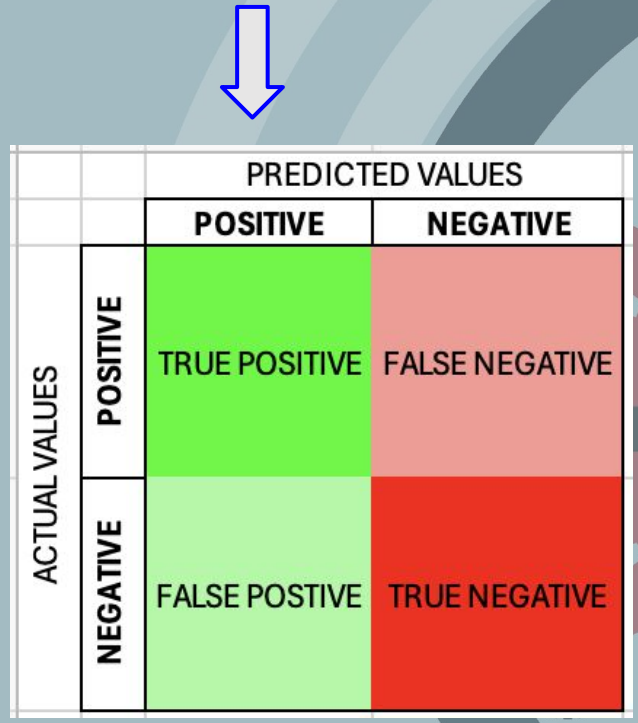
Training


Values
Time
Underfitted
Values
Time
Good Fit/Robust
Values
Time
Overfitted

Metrics

- Loss - defined by loss function
- Precision - **proportion of positive predictions that are correct**

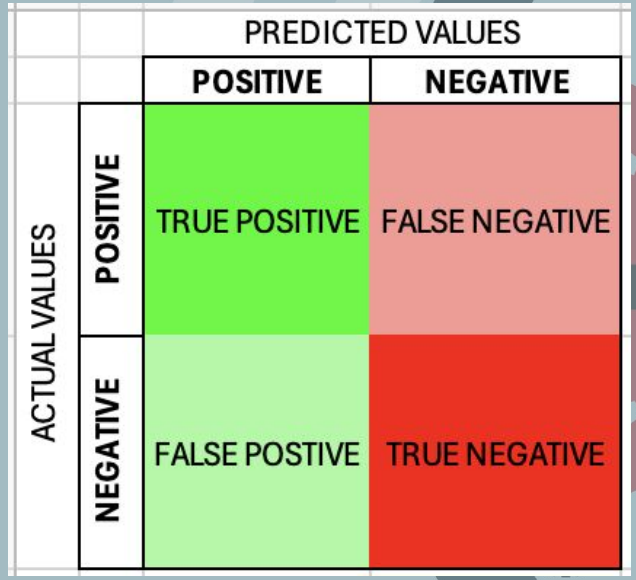
$$\text{Precision} = \frac{TP}{TP + FP}$$



Metrics

- Loss - defined by loss function
- Precision - proportion of positive predictions that are correct
- Recall - **proportion of actual positives correctly identified**

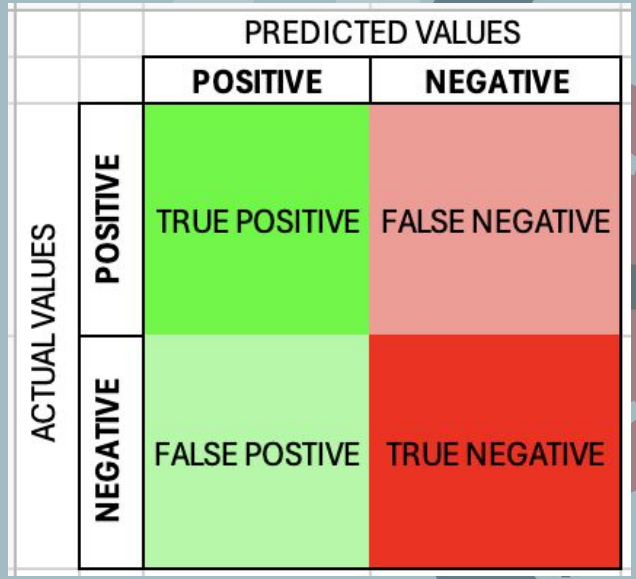
$$\text{Recall} = \frac{TP}{TP + FN}$$



Metrics

- Loss - defined by loss function
- Precision - proportion of positive predictions that are true
- Recall - proportion of actual positives correctly identified
- F-scores - **combination of precision and recall**

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

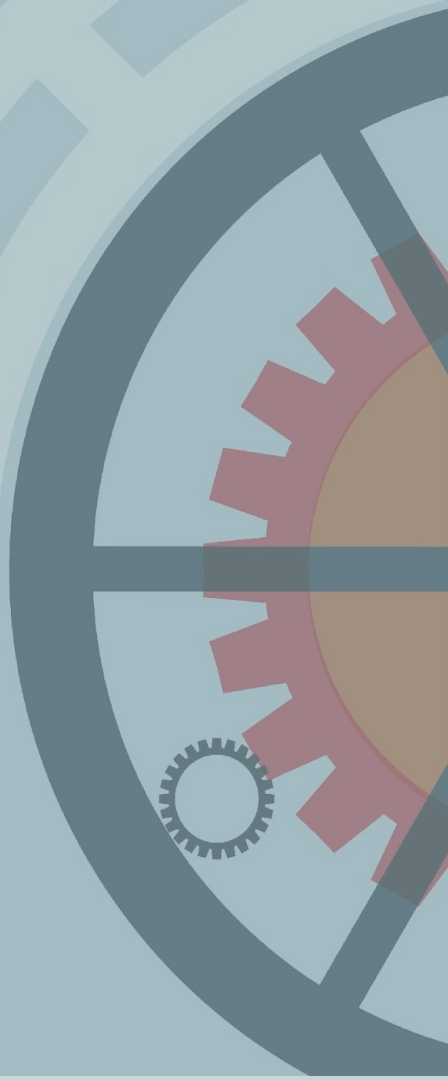


More on training CV models

Try these *Programming Historian* Lessons:

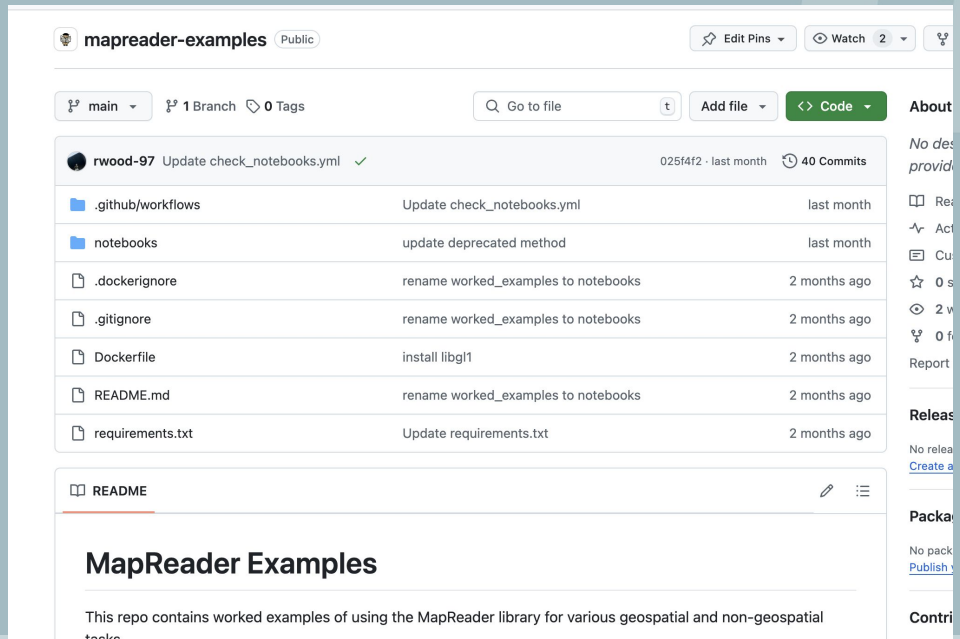
https://programminghistorian.org/en/lessons/computer-vision-deep-learning-pt1

https://programminghistorian.org/en/lessons/computer-vision-deep-learning-pt2



More MapReader example notebooks!

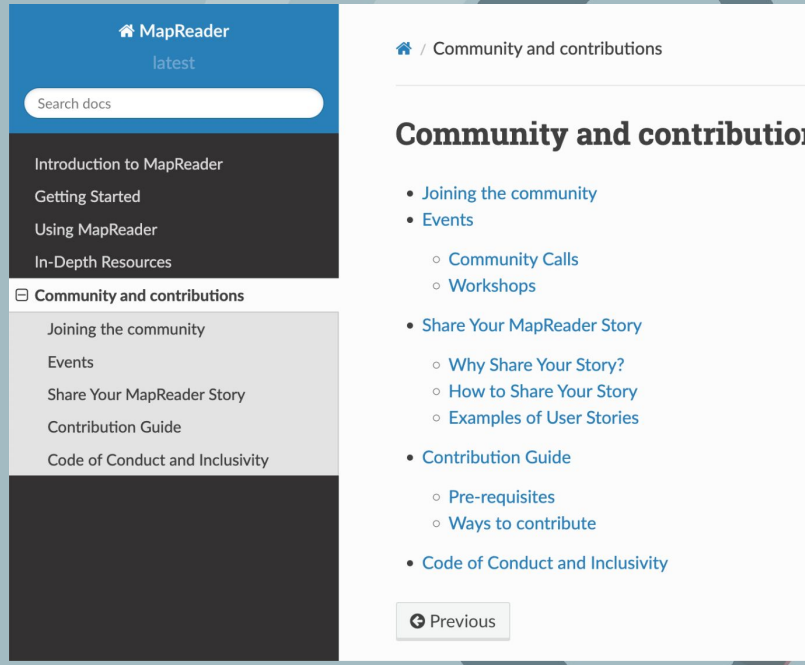
https://github.com/maps-as-data/MapReader-examples



Stay in touch: Community Calls

https://mapreader.readthedocs.io/en/latest/community-and-contributions/index.html

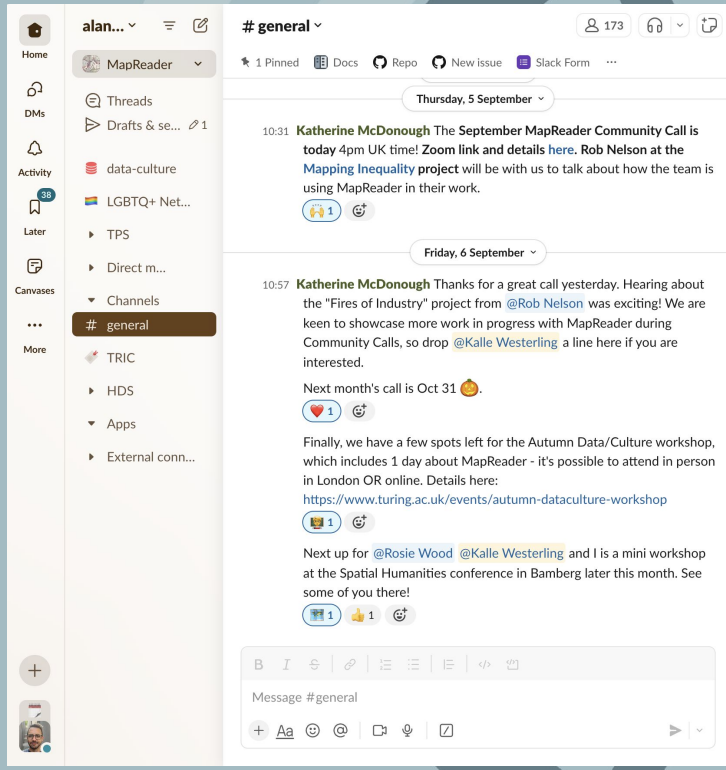
- **Connect with the Team:** Ask questions and get real-time updates from MapReader developers.
- **Learn Best Practices:** Pick up tips and workflows from experts and users alike.
- **Hear Success Stories:** Discover how others are using MapReader in their research.
- **Influence Development:** Share your ideas to help shape future features.
- **Build Community:** Network with fellow researchers and developers.



Stay in touch: Slack

[MapReader slack sign up form]

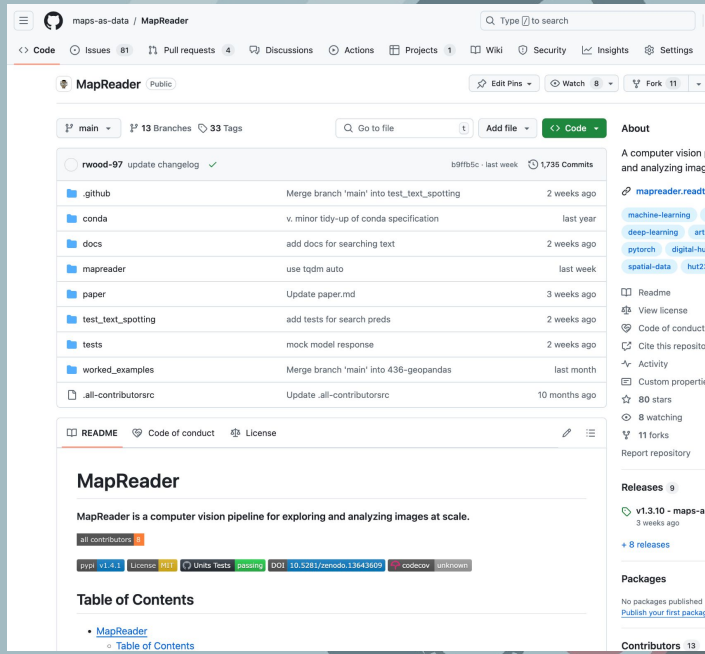
- **Get Help Fast:** Ask questions and get immediate support from the team and community.
- **Stay Informed:** Hear about new releases, features, and events first.
- **Share Ideas:** Discuss, learn, and exchange real-world use cases with others.
- **Shape the Future:** Contribute feedback, suggestions, or code to improve MapReader.
- **Connect with Experts:** Network with researchers and developers using MapReader.



Stay in touch: Github

https://github.com/maps-as-data/MapReader

- **Share Feedback:** Report bugs, suggest features, or give general input.
- **Improve Docs:** Help make guides and tutorials clearer by raising issues.
- **Track Progress:** Follow development updates in GitHub discussions.
- **Contribute Code:** Submit pull requests to add features or fix bugs. (For those who are more code-inclined!)
- **Collaborate:** Join discussions to shape the future of MapReader.



Thank you!