

Unlocking the Geospatial Past with Deep Learning – Establishing a Hub for Historical Map Data in Switzerland

Magnus Heitzler^{a,*}, Lorenz Hurni^a

^a Institute of Cartography and Geoinformation, ETH Zürich, Magnus Heitzler, hmagnus@ethz.ch, Lorenz Hurni, lhurni@ethz.ch

* Corresponding author

Keywords: Historical Maps, Artificial Neural Networks, Machine Learning, Spatial Data Infrastructures, Geoportals

Abstract:

Thoroughly prepared historical map data can facilitate research in a wide range of domains, including ecology and hydrology (e.g., for preservation and renaturation), urban planning and architecture (e.g., to analyse the settlement development), geology and insurance (e.g., to derive indicators of past natural hazards to estimate future events), and even linguistics (e.g., to explore the evolution of toponyms). Research groups in Switzerland have invested large amounts of time and money to manually derive features (e.g., pixel-based segmentations, vectorizations) from historical maps such as the Dufour Map Series (1845 – 1865) or the Siegfried Map Series (1872 – 1949). The results of these efforts typically cover limited areas of the respective map series and are tailored to specific research questions.

Recent research in automated data extraction from historical maps shows that Deep Learning (DL) methods based on Artificial Neural Networks (ANN) might significantly reduce this manual workload (Uhl et al. (2017), Heitzler et al. (2018)). Yet, efficiently exploiting DL methods to provide high-quality features requires detailed knowledge of the underlying mathematical concepts and software libraries, high-performance hardware to train models in a timely manner, and sufficient amounts of data.

Hence, a new initiative at the Institute of Cartography and Geoinformation (IKG) at ETH Zurich aims to establish a hub to systematically bundle the efforts of the many Swiss institutes working with historical map data and to provide the computational capabilities to efficiently extract the desired features from the vast collection of Swiss historical maps. This is primarily achieved by providing a spatial data infrastructure (SDI), which integrates a geoportal with a DL environment (see Figure 1).

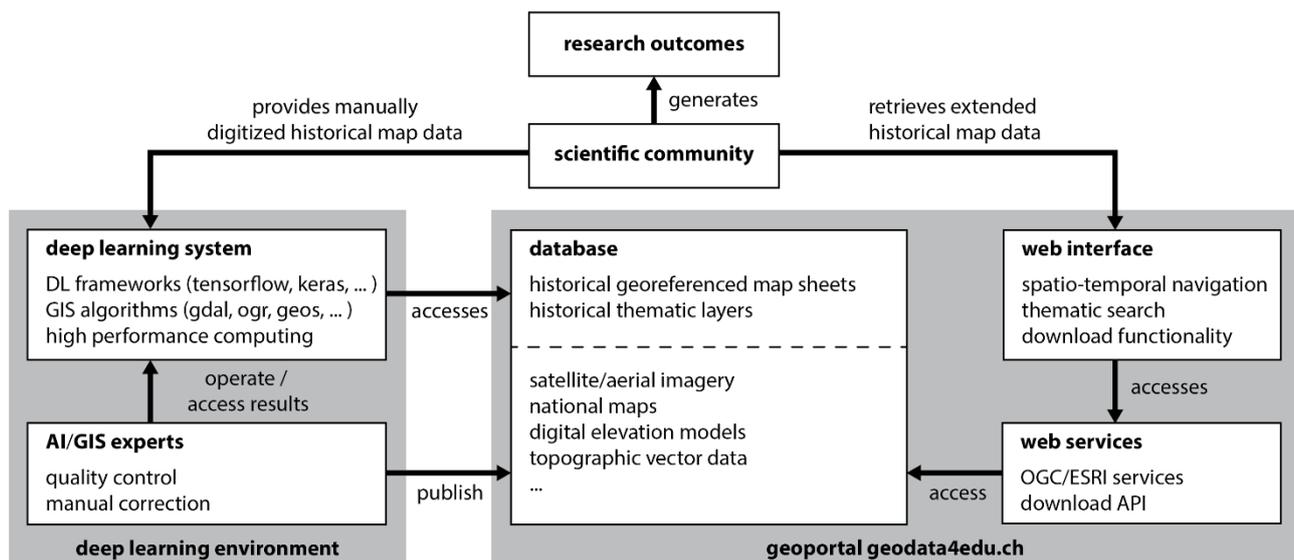


Figure 1. Spatial data infrastructure for the systematic preparation of Swiss historical maps for scientific purposes.

The SDI builds on top of the geoportal geodata4edu.ch (G4E), which was established to facilitate the access of federal and cantonal geodata to Swiss academic institutions. G4E inherently supports the integration and exploration of spatio-temporal data via an easy-to-use web interface and common web services and hence is an ideal choice to share historical map data. Making historical map data accessible in G4E is realized using state-of-the-art software libraries (e.g., Tensorflow, Keras), and suitable hardware (e.g., NVIDIA GPUs). Existing project data generated by the Swiss scientific community serve as the initial set to train a DL model for a specific thematic layer. If such data does not exist it is generated manually. Combining these data with georeferenced sheets of the corresponding map series allows the DL system to learn a way of obtaining the expected results based on the input map sheet. In the common case where an actual vectorization of a thematic layer is required, two steps are taken. First, the underlying ANN architecture yields a

segmentation of the map sheet to determine which pixel is part of the feature type of interest (e.g., by using a fully convolutional architecture such as U-Net (Ronneberger et al. (2015)) and, second, the resulting segmentations will be vectorized using GIS algorithms (e.g., using methods as described in Hori & Okazaki (1992)). These vectorizations undergo a quality check and might be directly published in G4E if the quality is considered high enough. In addition, the results may be manually corrected. A corrected dataset may have a greater value for the scientific community but might be time consuming to create. However, it has also the advantage to serve as additional training data for the DL system. This may lead to a positive feedback loop, which allows the ANN to gradually improve its predictions, which in turn improves the vectorization results and hence reduces the correction workload. Figure 2 shows automatically generated vectorizations of building footprints after two such iterations. Special emphasis was put on enforcing perpendicularity without requiring human intervention. At the time of writing, such building polygons have been generated for all Siegfried map sheets.

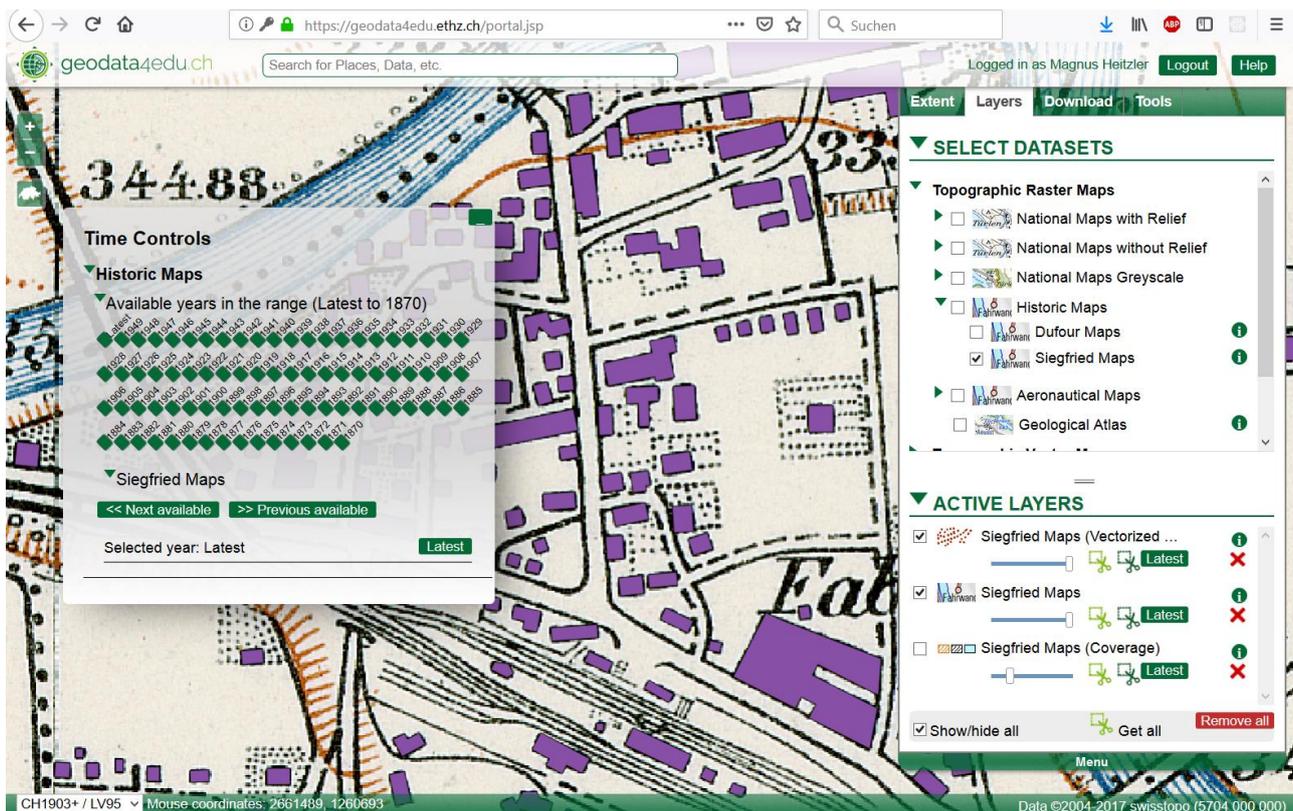


Figure 1. The geodata4edu.ch platform facilitates the download of automatically produced high-quality vectorizations of building footprints (purple) generated using deep learning and GIS methods.

It is worth emphasizing that showing the ability of generating high-quality features of single thematic layers at a large scale and making them easily available to the scientific community is a key aspect when establishing a hub for sharing historical map data. Research groups are more willing to share their data if they see that the coverage of the data they produce might get multiplied and if they realize that other groups are providing their data as well. Apart from the benefits for research groups using such data, such an environment also allows to facilitate the development of new methods to derive features from historical maps (e.g., for extraction, generalization). The current focus lies on the systematic preparation of all thematic layers of the main Swiss map series. Afterwards it is aimed to place higher emphasis on the fusion of the extracted layers. In the long-term, these efforts will lead to a comprehensive spatio-temporal database of high scientific value for the Swiss scientific community.

References:

- Heitzler, M., Gkonos, C., Tsorlini, A., & Hurni, L. (2018). A modular process to improve the georeferencing of the Siegfried map. *e-Perimtron*, 13(2), 85-100.
- Hori, O., & Okazaki, A. (1992). High Quality Vectorization Based on a Generic Object Model. In: Baird, H.S., Bunke H., Yamamoto, K. (eds.) *Structured Document Image Analysis*. Springer, Berlin, Heidelberg.
- Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, 234-241. Springer, Cham.
- Uhl, J., Leyk, S., Chiang, Y., Duan, W. & Knoblock, C. (2017). Extracting human settlement footprint from historical topographic map series using context-based machine learning. *8th International Conference of Pattern Recognition Systems (ICPRS 2017)*, Madrid, 2017, 1-6. doi: 10.1049/cp.2017.0144.