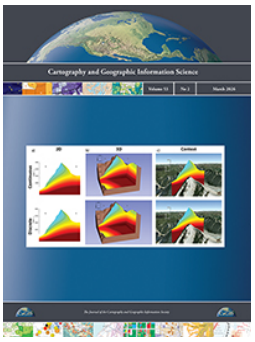

Taylor & Francis
Taylor & Francis Group

Cartography and Geographic Information Science

ISSN: 1523-0406 (Print) 1545-0465 (Online) Journal homepage: www.tandfonline.com/journals/tcag20

# Semantic segmentation of historical maps using Self-Constructing Graph Convolutional Networks

Lukas Arzoumanidis, Julius Knechtel, Jan-Henrik Haunert & Youness Dehbi

**To cite this article:** Lukas Arzoumanidis, Julius Knechtel, Jan-Henrik Haunert & Youness Dehbi (2026) Semantic segmentation of historical maps using Self-Constructing Graph Convolutional Networks, Cartography and Geographic Information Science, 53:2, 177-187, DOI: 10.1080/15230406.2025.2468304

**To link to this article:** https://doi.org/10.1080/15230406.2025.2468304



Published online: 13 Mar 2025.

CARTOGRAPHY AND GEOGRAPHIC INFORMATION SCIENCE
2026, VOL. 53, NO. 2, 177–187
https://doi.org/10.1080/15230406.2025.2468304

Taylor & Francis
Taylor & Francis Group

OPEN ACCESS   Check for updates

# Semantic segmentation of historical maps using Self-Constructing Graph Convolutional Networks

Lukas Arzoumanidis [a], Julius Knechtel [b], Jan-Henrik Haunert [b] and Youness Dehbi [a]

[a]Computational Methods Lab, HafenCity University Hamburg, Germany; [b]Institute of Geodesy and Geoinformation, University of Bonn, Germany

**ABSTRACT**
Historical maps represent an invaluable memory which should be preserved. Such kind of maps are, however, mostly scanned and stored as raster graphics which do not contain semantic information in a machine-readable form. To achieve a machine-readable state, an often expensive human intervention is needed in a fully manual or semi-automatic fashion. An automatic interpretation and a feature extraction is then inevitable for a map digitization and vectorization. Automatic approaches showed more and more convincing and promising results on challenging map corpora avoiding human interaction. This paper deals with the semantic segmentation of historical maps based on Graph Convolutional Networks (GCNs) to capture long-range dependencies between image features. This allows for an extension of the receptive field of Convolutional Neural Networks (CNNs) restricted on local dependencies. A Self-Constructing Graph (SCG) module has been applied to automatically induce the structure of the GCN. We performed experiments revealing promising results where our approach achieved an Mean Intersection over Union (mIoU) of 0.68, outperforming a state-of-the-art CNN dedicated to the semantic segmentation of historical maps.

**KEY POLICY HIGHLIGHTS**

- Automatic semantic segmentation of heterogeneous historical map corpora
- Combination of Convolutional Neural Network (CNN) and Graph Convolutional Network (GCN) to capture long-range dependencies
- Automatic induction of the graph structure of GCN using Self-Constructing Graph module
- Demonstrated superiority over state-of-the-art CNN-based methods

**ARTICLE HISTORY**
Received 15 May 2024
Accepted 4 February 2025

**KEYWORDS**
Historical map processing; semantic segmentation; Graph Convolutional Networks; Self-Constructing Graph; heterogeneous corpora

## 1. Introduction

Historical maps precisely capture the spatial reality over extended periods of time (Komac & Gašperič, 2023). They are a fundamental source of geographic information (Chiang et al., 2020; Komac & Gašperič, 2023), e.g. as a prerequisite for backtracking and monitoring environmental changes including land use change (Mäyrä et al., 2023), urbanization (Budig et al., 2016; Uhl et al., 2021, 2022) and landscape ecology (Auffret et al., 2017). To avoid a deterioration of this information and consequently preserve and make it widely available, often a *digitization* is performed, which mostly corresponds to the process of scanning a historical document and storing the resulting raster graphic in a database. However, in this case in this case, the scanned historical documents, i.e. maps, neither have searchable metadata nor are semantically enriched (Chiang et al., 2020; Li et al., 2020). Hence, invaluable spatial information remains unreachable. Thus, the ultimate goal is to digitize historical maps in a way that their information and content is readable, searchable and analyzable by machines. In combination with a database, this allows for an effective and intuitive analysis and comparison with current map features (Schlegel, 2019; Shbita et al., 2023).

As yet, the digitization is often performed in a manual, or at best, semi-automatic fashion which is time-consuming and labor-intensive (Li et al., 2020). Additionally, a low level of consistency might result from different people involved in the process of manual digitization. Therefore, an increase in the level of automation is deemed desirable. However, the complexity arising from the extensive stylistic diversity among historical maps, including their unique typefaces and semiology, presents significant challenges for the development of automated approaches. Additionally, data-dependent uncertainties – such as image processing difficulties caused by the deterioration of original

**CONTACT** Lukas Arzoumanidis lukas.arzoumanidis@hcu-hamburg.de

map sheets (e.g. paper distortion or color fading) – further compound these challenges. However, the great success of machine learning-based approaches, especially deep neural networks, for the semantic segmentation of challenging datasets, e.g. architectural floorplans (Knechtel et al., 2024), motivated the investigation of new approaches for the semantic segmentation of historical maps and the identification and extraction of historical map features (Petitpierre et al., 2021; Wu et al., 2023).

Given the extensive amount of scanned historical maps and the increasing importance of image segmentation for successful information extraction, we further investigate and demonstrate the potential of automatic semantic segmentation for the digitization of complex heterogeneous map corpora by applying deep neural networks. Inherent to the learning process, however, deep neural networks need an extensive amount of allocated memory. In this context, the input size of learning examples and the depth of the network architecture as well as the according operations such as convolutions or poolings are decisive. To overcome these limitations, maps are usually cropped into smaller image patches, effectively reducing contextual information in some map regions. This increases, however, the grade of data-dependent uncertainty. In recent years, Convolutional Neural Networks (CNNs) were predominantly used for tasks of semantic segmentation. However, this kind of networks does not explicitly consider the long-range dependency between learned latent features, due to the intrinsic locality of the convolution (Wu et al., 2023). In this context, convoluting in the latent feature space rather than on fixed map regions of full-resolution images may reduce a network's computational costs. Thus, the main contribution of this article is to apply a deep learning-based approach for the semantic segmentation of historical maps that uses a Self-Constructing Graph Convolutional Network (SCGCN) introduced by Liu et al. (2021) to not only capture local contextual information but also the long-range dependencies. The latter are considered between a-priori extracted and learned features rather than targeting individual pixels within each input map. Figure 1 summarizes the main idea of our approach. Based on a set of historical maps and their corresponding semantically annotated masks, a Self-Constructing Graph Convolutional Network has been trained. In a test phase, our method is able to predict pre-defined semantic classes for unseen historical maps. More details on the architecture of the network used will be shown in Section 3. To provide evidence that graph convolutional networks are a viable option for the semantic segmentation of historical maps, we compare our approach against the CNN-based approach of Petitpierre et al. (2021).

The remainder of this paper is structured as follows: Section 2 gives insights into recent work related to the digitization of historical maps. Section 3 briefly underlines the methodological background of the applied method. In Section 4, a comparative study is performed to demonstrate the superiority of our approach against the segmentation method of Petitpierre et al. (2021). The paper is then summarized in Section 5 and also providing an outlook for future research.

## 2. Related work

In the last three decades, several semi-automatic and automatic approaches to digitize historical maps have been proposed (Jelena et al., 2018). The digitization of such maps usually requires visual object and pattern recognition, e.g. to identify edges of building footprints, land use classes or the structure and meaning of symbols

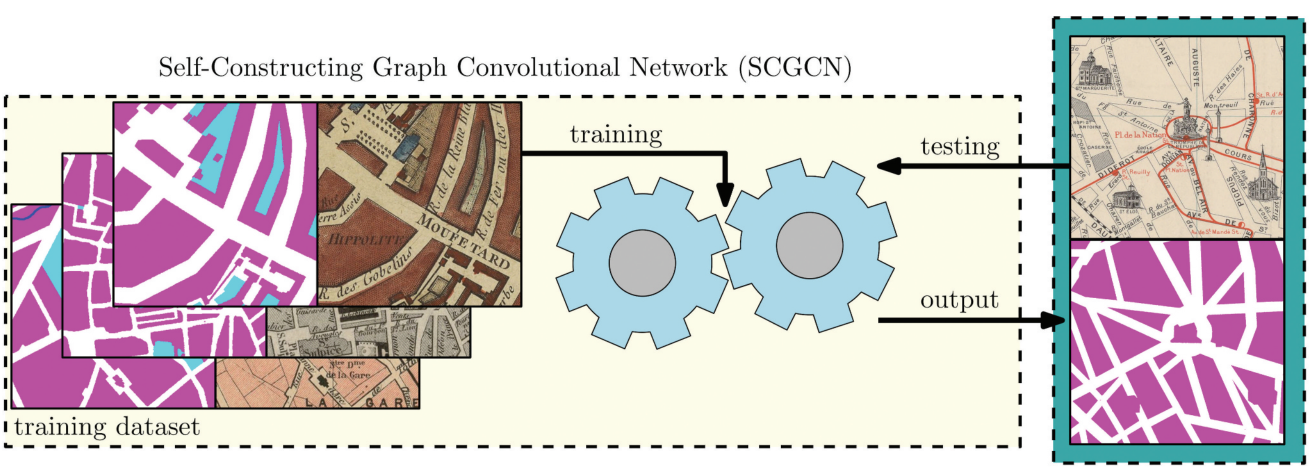


Figure 1. Overview of our approach for the semantic segmentation of historical maps using a Self-Constructing Graph Convolutional Network (SCGCN).

and the characters of text labels (Chiang et al., 2020). Semi-automatic approaches for analyzing historical maps were the first attempt to replace tedious and time-consuming manual digitization. In the past, semi-automatic approaches have been proposed for different tasks, such as the extraction of building footprints and administrative boundaries (Chrysovalantis & Tziokas, 2020; Gede et al., 2020; le Riche, 2020; Piechl, 2020), or the extraction of road networks (Chiang & Knoblock, 2012, 2013; Jiao et al., 2021). However, since semi-automatic approaches heavily rely on manual user input, they do not scale well for processing larger map corpora with various types of historical maps and heterogeneous contents or poor image quality due to deterioration effects (Chiang et al., 2020).

A special challenge when automating the digitization of historical maps is the extraction of labels and text. This holds particularly when labels vary strongly in shape, size, orientation or type and overlap each other as they convey valuable information (Chiang et al., 2020). Contextual data include the precise location and the characteristics of a map object and can thereby help to correct recognition and georeferencing errors (Li et al., 2021; Schlegel, 2023). Recently, machine learning techniques focused on automated text recognition and extraction (Laumer et al., 2020; Schlegel, 2021; Weinman, 2013). A major barrier to effectively digitize historical maps poses the extraction of geographic locations. Frequently, the location is determined through a slow and tedious manual georeferencing (Bahgat & Runfola, 2021). However, recent approaches to automate georeferencing for historical maps are related to the previously mentioned approaches for extracting text labels (Bahgat & Runfola, 2021; Luft, 2020; Milleville et al., 2020). This is, however, beyond the scope of our paper. In this article, the focus is shifted toward a new approach to the semantic segmentation of historical maps, paving the way to vectorization in a subsequent step.

Due to the huge development and high increase of computational capacities, most automatic approaches nowadays benefit from the application of deep learning. Hence, machine learning approaches incorporating deep artificial structures have been developed, showing state-of-the-art results in many different research fields including the semantic segmentation of historical maps and feature extraction. However, the overall complexity and diversity in historical maps, i.e. their extensive stylistic variety and their individual typefaces as well as their deterioration, still pose a major difficulty to develop automatic approaches since an extensive amount of training data is needed (Li, 2019; Li et al., 2021; Nobajas, 2020; Pack & Kim, 2023; Sobotkova et al., 2023). As semi-automatic approaches do not generalize well enough on heterogeneous corpora, automatic approaches being independent from human interaction are desirable. In this context, deep learning promises a large potential for the automatic digitization of historical maps (Chiang et al., 2020).

Deep learning has been already applied on historical maps, e.g. for the extraction of wetlands (Jiao et al., 2020; Wu et al., 2022) or detection of land use change (Mäyrä et al., 2023), for the extraction of archeological features from historical map series (Garcia-Molsosa et al., 2021) or for the extraction of surface mine extents (Maxwell et al., 2020) as well as for the extraction of road networks (Avcı et al., 2022; Ekim et al., 2021; Jiao et al., 2022a, 2022b; Uhl et al., 2022) and building footprints (Heitzler & Hurni, 2020). Furthermore, urban planners can retrieve extensive amounts of information from historical maps, such as building footprints or parcel marks, by applying deep learning techniques to automatically extract cadastral maps (Chen et al., 2021; Jelena et al., 2018) as well as urban zoning plans (Guo et al., 2018; Uhl et al., 2021).

Ares Oliveira et al. (2018) developed a slightly modified U-Net architecture that can be used for the semantic segmentation of historical documents. Their network uses a pretrained *ResNet-50* backbone and shows sophisticated results for the semantic segmentation of historical documents. Following their idea, Petitpierre et al. (2021) propose an automatic approach for the semantic segmentation of historical maps using the network previously proposed by Ares Oliveira et al. (2018) yielding strong results. The dataset used in their work can be characterized as a heterogeneous corpus consisting of historical maps dating from different epochs and having various scale and figuration of Paris (Petitpierre et al., 2021). This dataset will be subject for comparing the CNN presented by Ares Oliveira et al. (2018) with the adopted GCN in this work.

Recently, deep learning-based techniques have emerged that aim to leverage long-range dependencies to exploit the contextual information in maps. Wu et al. (2023) propose a U-Net-based network that encodes spatio-temporal features and integrates cross-attention transformers to aggregate information over a larger spatial range, thereby providing spatio-temporal context to the transposed convolutions in the decoder. In another study, Wu et al., (2022) address the noise inherent in historical maps by employing Bayesian deep learning for the semantic segmentation of hydrological features. Their approach integrates atrous spatial pyramid pooling (ASPP) into a U-Net-based network and adds uncertainty estimation as an auxiliary task to incorporate multi-scale contextual information.

To further emphasize the integration of long-range feature dependencies and spatial context into deep neural networks, and building upon the ideas of Liu et al. (2020), our approach employs and adapts Self-Constructing Graphs (SCG) for the automatic semantic interpretation of historical maps. SCGs have been used for the semantic labeling of urban scenes based on the publicly available ISPRS Potsdam and Vaihingen dataset (Liu et al., 2021) and for the segmentation of airborne images from the agriculture-vision challenge dataset later on (Liu et al., 2020). Recently, Knechtel et al. (2024) applied SCGs as part of a multi-task network for the semantic interpretation of architectural floorplans. Instead of a tessellation of room geometries with different categories in building floorplans, our approach investigates the transferability and suitability of SCGs to address the interpretation of historical maps as segmentations of areas of different object classes.

## 3. Methodology

For the semantic segmentation of the historical maps, we apply a network first published by Liu et al. (2021), which among others utilizes the concept of Graph Convolutional Networks (GCNs). Applying a GCN allows information from spatially distant pixels to be processed with a shallow network. In comparison, the use of traditional Convolutional Neural Networks (CNNs) requires a deep network, i.e. multiple layers, to get a comparable receptive field, which among other things leads to high memory requirements.

The network itself, which is depicted in Figure 2, consists of three interlinked parts: An encoder, the so-called Self-Constructing Graph Module (SCG) and a GCN. The encoder is a state-of-the-art CNN extracting features from the input image. In our work, we use a ResNet-101 encoder, which in contrast to Liu et al. (2021) consists of only three layers. The features, however, cannot be directly used as input for the GCN since it requires a graph structure. Without any additional expert knowledge, this structure is automatically retrieved using the Self-Constructing Graph Module. This module again consists of two components, an encoder part (ENC) and a decoder part (DEC). Based on Gaussian Parameters $\mu$ and $\sigma$ a new embedding for the features is learned in ENC while utilizing a Kullback–Leibler divergence as loss function $\mathcal{L}_{\text{kl}}$. Subsequently, in DEC, the adjacency Matrix $A$ is computed, which establishes the graph structure needed for the GCN. An edge between two corresponding nodes, i.e. features, in the graph is introduced based on their respective similarity. Hence, the diagonal of $A$, which represents the similarity of the feature to itself, should be close to one, leading to a diagonal logarithmic regularization as loss function ($\mathcal{L}_{\text{dl}}$) for the decoder part. The resulting graph with its nodes and edges is then used as input for two consecutive GCN layers. By performing convolutions on the graph structure, new embeddings for the nodes are learned on which the allocation to the individual classes takes place. As loss function for the semantic segmentation an *Adaptive Class Weighting Loss* (ACW loss) is applied (Liu et al., 2021). The classes in the dataset are highly imbalanced, i.e. different in terms of their frequency of occurrence. Nevertheless, the ACW loss is not using weights that were precomputed based on the whole dataset, but performs an iterative batch-wise class rectification. The entire loss function of the network can therefore be summarized as:

$$\mathcal{L} = \mathcal{L}_{\text{kl}} - \mathcal{L}_{\text{dl}} + \mathcal{L}_{\mathcal{ACW}}. \quad (1)$$

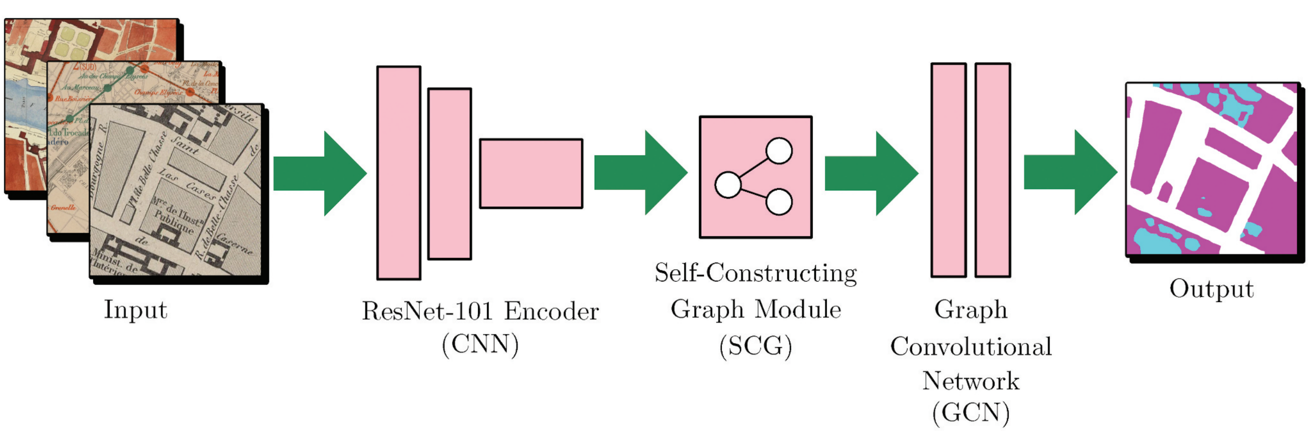


Figure 2. Overview of the used network. A batch of historical maps is given as input. A CNN consisting of a ResNet-101 encodes features into feature maps, which are subsequently translated into a graph representation by the Self-Constructing Module allowing for the prediction of historical maps as output.

For more information on both the network architecture and the loss function, the interested reader is referred to Liu et al. (2021).

## 4. Experimental results

In order to investigate the suitability of GCNs for the semantic segmentation of historical maps, our experiments have been performed on the above-mentioned dataset from Paris which is openly available in Zenodo (Petitpierre, 2021). The dataset consists of maps dating from 1800 to 1950 at a scale between 1:2000 and 1:25000 and stems from the *Bibliothéque nationale de France* and the *Bibliothéque historique de la Ville de Paris* (Petitpierre et al., 2021). Illustrating the differences in scale, Figure 3b) represents a map with a larger scale, while the map in Figure 3d) has a smaller scale. The underlying maps are culturally and geographically homogeneous as they stem from Paris. However, they are figuratively very diverse expressed through the usage of different styles, e.g. fonts, colors, symbols, and the provided information. In particular, information not only concerning mobility, such as the road or underground network (as highlighted in Figure 3c), but also on the water system, the catacombs, the administrative divisions (as can be seen in Figure 3a), or the extent of urban structures (as captured in Figure 3d)–3f) are present. Regarding the deterioration of the original map sheets, Figure 3e) indicates paper distortion through creases that resulted from folding the map. Another common issue with historical maps is color fading as can be seen in the left third of Figure 3a) or slightly above the center of Figure 3b). Additionally, mark ups or impurities as indicated in Figure 3e) and mistakes during scanning, e.g. a low image resolution (noticeable in Figure 3d) or a captured map strap (as can be seen in Figure 3f) add more complexity to the processing of this corpus.

**Figure 3.** Heterogeneity of the Paris dataset.

The dataset provides five different classes, i.e. *non-built*, *building blocks*, *road network*, *water* and *frame*. It originally consisted of 300 images for training and 30 for validation and testing with a resolution of $1000 \times 1000$ pixels, but for our training we cropped the images in patches of $500 \times 500$ pixels. We used a pretrained *ResNet-101* as encoder, $32 \times 32$ nodes for training, $64 \times 64$ nodes for the inference, a batch size of 9 and applied a learning rate of $3.47 \times 10^{-4}$. All experiments have been conducted on a Nvidia RTX 3080 with 10GB of DDR6X VRAM with *Cuda 11* libraries. *Pytorch 1.8* and *Python 3.8* have been used to modify and run both networks.

Figure 4 and Table 1 provide quantitative results of our approach in comparison with the results from *dhSegment* (Petitpierre et al., 2021). In their method, the *frame* class was predicted during a pre-segmentation as part of a preprocessing step, and their network was subsequently trained on the remaining four classes: water, non-built area, road network, and building block. In contrast, our network was trained simultaneously on all five classes available in the dataset, including the *frame* class. Still, our network outperforms *dhSegment* considering the mean values of all tested quality measures, i.e. IntersectionOverUnion (IoU), Precision, Recall and $F_1$ score. Our approach achieved an mIoU of 0.68, which is 5% higher than the mIoU of 0.63 achieved by *dhSegment*. Furthermore, our approach results in a Precision of 0.78 while *dhSegment* leads to a Precision of 0.72 indicating a difference of 6%. Lastly, the Recall of our approach is 0.85 while the Recall of *dhSegment* is 0.81 demonstrating a small lead in performance overall. By balancing precision and recall, our approach achieves an $F_1$ score

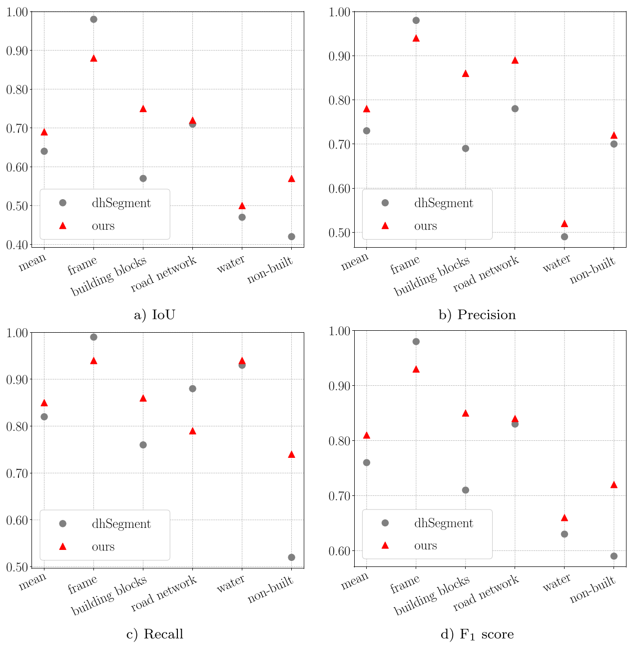


**Figure 4.** Different quantitative measures to compare results from our approach and dhSegment. Abbreviations: building blocks as "b. blocks,"non-built as "n.-built," and road network as "road net".


**Figure 5.** The resulting confusion matrix for our approach. Abbreviations: building blocks as "b. blocks,"non-built as "n.-built," and road network as "road net".

of 0.81, compared to 0.76 for *dhSegment*, demonstrating an improvement of 5%. Although the pre-trained frame prediction performs better, our network shows significant improvement in the other classes considering all four measures, e.g. of 0.19 in IoU, 0.18 in Precision and 0.14 in $F_1$ score of building blocks and of 0.15 in IoU, even 0.23 in Recall and 0.13 in $F_1$ score for non-built structures. While the Recall for the road network class is slightly lower, our approach demonstrates a 0.11 improvement in Precision, with IoU results also showing a slight advantage. Additionally, our method achieved a 4% higher IoU and a 3% higher Precision for the *water* class compared to dhSegment, reflecting overall improved outcomes, as illustrated in Figure 4 and Table 1.

The confusion matrix of our results, shown in Figure 5, indicates that the road network is frequently misclassified as building blocks. Additionally, according to the confusion matrix, our approach frequently confuses water bodies with non-built area while non-built area is often confused with building blocks as depicted in Figure 5. In contrast to our results, *dhSegment* frequently misclassifies building blocks as non-built areas and vice versa. Unfortunately, the confusion matrix entries from Petitpierre et al. (2021) are not normalized, complicating direct comparison. Additionally, our approach can predict the frame, whereas *dhSegment* is not designed to handle frame prediction; in their approach, the *frame* class is incorporated in a pre-segmentation as part of a preprocessing step. Interestingly, the confusion matrix reveals that the accuracy scores for the road network and water bodies are of

**Table 1.** Per-class and mean percentage values of IoU, precision, recall and $F_1$ score using our SCGCN approach and *dhSegment* (dhS.) for the classes: frame, water, non-built area, road network, and building block.

| | *IoU* | | *Precision* | | *Recall* | | *$F_1$ score* | |
|---|---|---|---|---|---|---|---|---|
| | *ours* | *dhS.* | *ours* | *dhS.* | *ours* | *dhS.* | *ours* | *dhS.* |
| *mean* | **0.68** | 0.63 | **0.78** | 0.72 | **0.85** | 0.81 | **0.81** | 0.76 |
| *frame* | 0.88 | **0.98*** | 0.93 | **0.98*** | 0.93 | **0.99*** | 0.93 | **0.98*** |
| *building block* | **0.75** | 0.56 | **0.86** | 0.68 | **0.85** | 0.76 | **0.85** | 0.71 |
| *road network* | 0.72 | 0.71 | **0.89** | 0.78 | 0.79 | **0.88** | 0.84 | 0.83 |
| *water* | **0.50** | 0.46 | **0.51** | 0.48 | 0.93 | 0.92 | 0.66 | 0.63 |
| *non-built* | **0.57** | 0.42 | 0.71 | 0.70 | **0.74** | 0.51 | **0.72** | 0.59 |

Compared to our automatically inferred values, the quantitative measures for the class *frame* in the approach of Petitpierre et al. (2021) (indicated by an asterisk) were achieved in a pre-segmentation as part of a preprocessing step. Values highlighted in bold indicate larger differences in performance. All values are rounded to two decimal places.

a similar range. However, as reported in Table 1, the IoU for the road network is significantly higher. This discrepancy may arise from the fact that the IoU evaluates regions – in this case, masks – whereas the accuracy metric in the confusion matrix is pixel-based by comparison with the ground truth. While our network appears to predict regions with significant overlap, it may misclassify boundaries, leading to an improved IoU but a reduced accuracy due to incorrect boundary predictions. As illustrated in Figure 3, roads in this specific map corpus are characterized by finer boundaries and more complex geometries compared to water bodies, potentially amplifying this discrepancy.

Figure 6 depicts exemplary input images a) &; d) &; g), their respective ground truth b) &; e) &; h) and the corresponding result of our network c) &; f) &; i). The

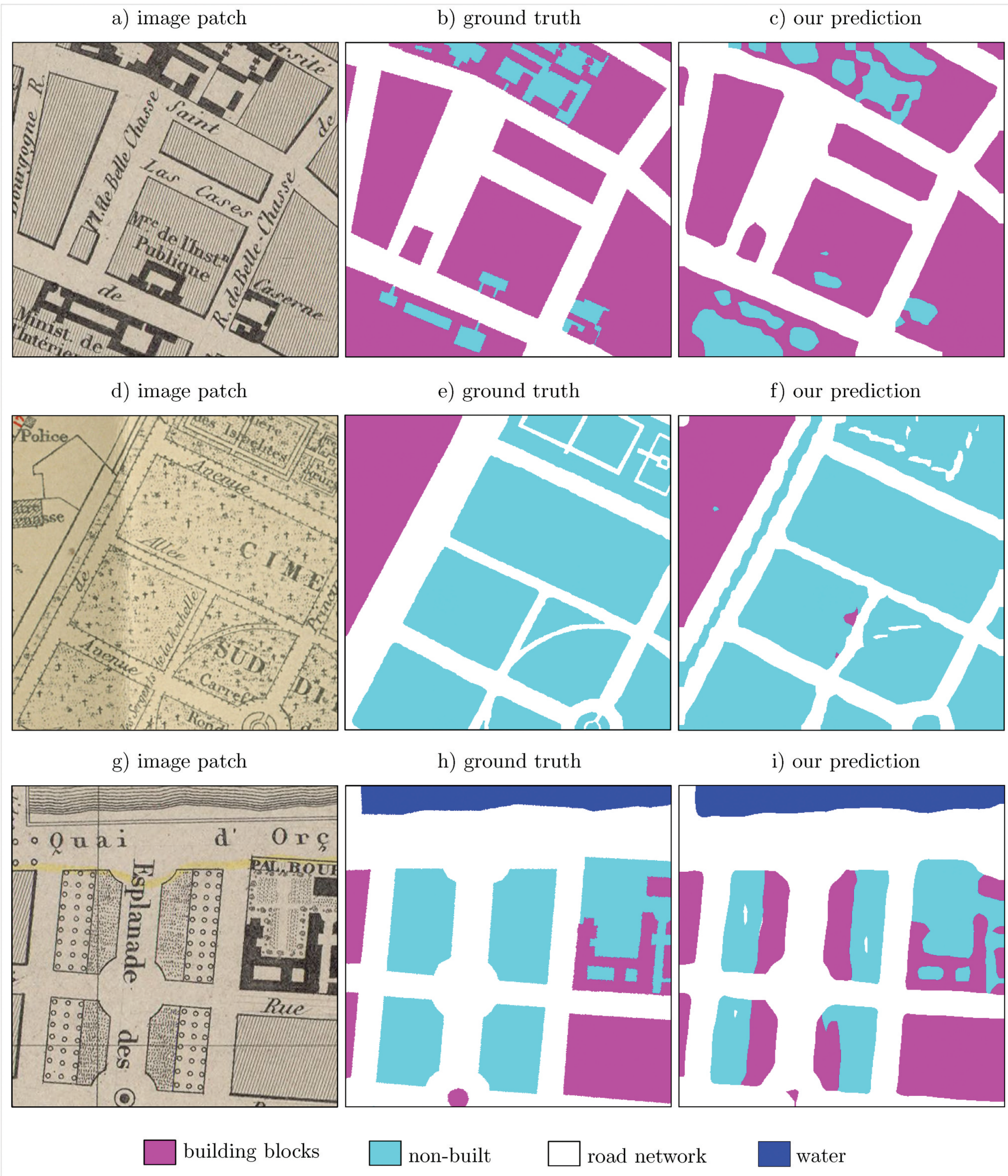


**Figure 6.** Exemplary qualitative results of our approach.

prediction of larger and simpler geometries, like the building blocks in Figure 6c) or the non-built area in Figure 6f) or the water in Figure 6i) is perceived as accurate. Our method is even able to reveal annotation errors, e.g. the non-built stripe in the main road in Figure 6d) missing in Figure 6e) of the ground truth. Finer structures, i.e. more complex geometries, like the non-built area in Figure 6c) & 6i) are insufficiently predicted. Basically, our approach misclassifies the non-built areas in the middle of Figure 6i). However, a closer look at the according image patch in Figure 6g) reveals that the SCGCN was able to differentiate between the symbology of different possibly underlying vegetation types.

The improved performance of GCNs might be attributed to their ability to incorporate long-range dependencies between individual features without relying on deep structures with a high number of layers as a traditional CNN (Liu et al., 2021). This assumption is supported by the observation that other map elements, e.g. grid lines, labels or sketches, describing points of interest or streets, do not prevent the model from accurate predictions as can be seen in Figure 6. An exhaustive assessment of the impact of the graph structure in our method and its impact on the quality of results is, however, beyond the scope of this article and will be subject of upcoming research.

As previously mentioned, our approach struggles to reliably segment water, as can be seen in Figure 5. Approximately 40% of pixels belonging to water bodies are misclassified as non-built areas. While these two classes are generally assumed to be distinct, a closer examination of the historical maps reveals a high degree of visual similarity between them, particularly in the hachures and colors used. Additionally, these classes are often spatially adjacent, which further contributes to the misclassification challenge.

Figure 7 illustrates two examples where similarities in hatching patterns and colors between the *water* and *non-built* area classes might prevent the network from

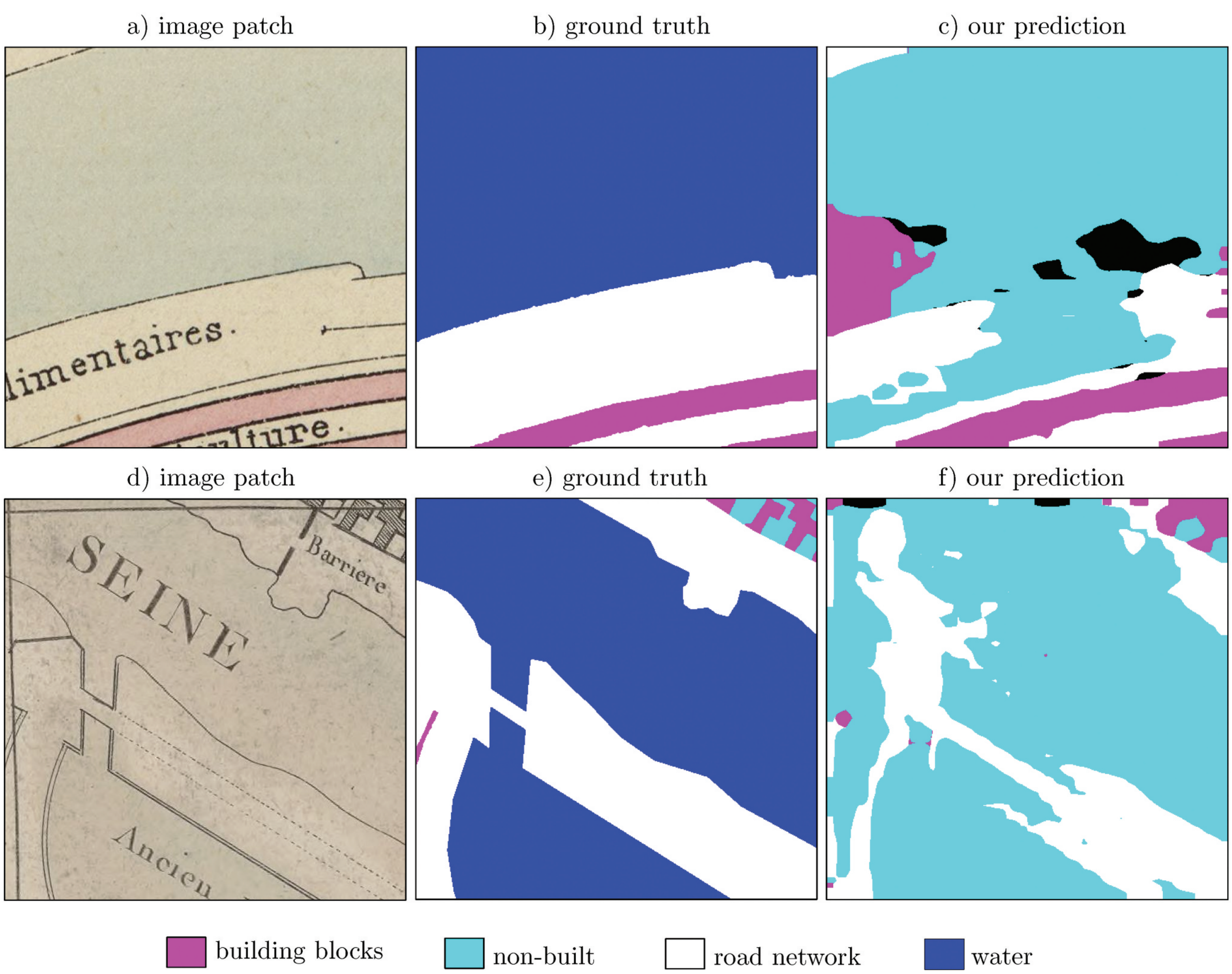


**Figure 7.** Exemplary qualitative results from our approach highlight instances of unsuccessful segmentation for water and non-built areas. The challenge primarily arises from the high similarity in color, hachures and shading patterns between water and non-built regions, which complicates accurate discrimination. The close resemblance in visual features can lead to overlapping classifications, underscoring the complexity of distinguishing these classes effectively.

accurately segmenting the maps. In the first instance (Figure 7a)–7c), the two classes show only slight color differences. However, the color representing water – a light or pastel green extending from the center to the top and spanning left to right – also frequently denotes non-built areas in other maps within the same corpus. This overlap introduces ambiguity, making it difficult for the network to differentiate these classes effectively. Such color overlap can cause confusion during training, further increasing the complexity of accurate segmentation.

In the second example, shown in Figure 7d)–7f), the water – specifically the Seine River – and non-built areas in other maps within the same corpus share nearly identical colors, introducing ambiguity and potentially causing confusion in the network during training. Additionally, the water in this instance has a slight green tint, visible around the "Seine" label and in the river bay below the "Barriere" label. This tint may be due to an intentional hatching pattern added by the map's creator, which has faded over time due to poor preservation, or it may result from dirt accumulation on the map, also due to inadequate preservation. This illustrates the inherent challenges of working with historical maps. Interestingly, although the water and street colors in this example are nearly identical with barely noticeable differences, the network is at least somewhat capable of distinguishing between the two classes. Overall, the frequent misclassification of water and non-built areas can likely be attributed to the heterogeneous nature of the historical map corpus used for training, where these classes sometimes share the same color or hatching within individual maps. This underscores the challenge of performing semantic segmentation on such diverse maps, particularly for certain classes.

## 5. Conclusion and outlook

Engineering an automatic procedure to extract valuable spatial information is obligatory for the digitization of historical maps. In this paper, we applied an approach that is able to identify four different land use classes and the map frame with a strong figurative diversity. Based on the comparative study, our approach using Graph Convolutional Networks for the semantic segmentation of historical maps outperforms the CNN-based state-of-the-art *dhSegment*. Our approach achieved an mIoU of 0.68, which is almost 5% higher than the mIoU of 0.63 achieved by *dhSegment*. Furthermore, it represents a promising method for the automatic detection and interpretation of historical map features. While our approach is already able to predict larger map objects with simpler geometries, the prediction of finer, more complex geometries needs to be further improved. Additional preprocessing and augmentation of training data and newer versions of feature encoders such as EfficientNet or encoders based on vision transformer models will be subject of future work to hopefully further improve the performance on heterogeneous datasets. This will pave the way for a subsequent vectorization and further automate the digitization of historical maps. Another aspect to address is the vectorization of the semantically segmented maps. In this context, the suitability of methods like Mixed Integer Linear Programming (MILP) to enforce boundary constraints between different classes will be topic of investigation. Another task worth of investigation is to perform a deeper analysis of the impact of the automatically generated Self-Constructing Graph (SCG). This will be addressed in an ongoing project dealing with the explainability and visualization inherent of deep structures.


## Acknowledgments

We gratefully acknowledge the open dataset of annotated historical maps provided by (Petitpierre, 2021) that has been used to compare our presented approach against theirs. Besides, we would like to express our gratitude to the open code from Liu et al. (2021).

## Disclosure statement

No potential conflict of interest was reported by the author(s).



## ORCID

Lukas Arzoumanidis http://orcid.org/0000-0001-6668-1695
Julius Knechtel http://orcid.org/0009-0000-8550-9700
Jan-Henrik Haunert http://orcid.org/0000-0001-8005-943X
Youness Dehbi http://orcid.org/0000-0003-0133-4099


## Data availability statement

Our code and pre-trained models are accessible at https://github.com/hcu-cml/SCGCN-histMap-segmentation.

## References


Ares Oliveira, S., Seguin, B., & Kaplan, F. (2018). dhSegment: A generic deep-learning approach for document segmentation. In *2018 16th International Conference on Frontiers in Handwriting Recognition (ICFHR)*, Niagara Falls, NY, USA (pp. 7–12). https://doi.org/10.1109/ICFHR-2018.2018.00011

Auffret, A. G., Kimberley, A., Plue, J., Skånes, H., Jakobsson, S., Waldén, E., Wood, H., Bullock, J. M., Cousins, S. A. O., Gartz, M., Hooftman, D. A. P., & Tränk, L. (2017). Histmapr: Rapid digitization of historical land-use maps in R. *Methods in Ecology and Evolution*, *8* (11), 1453–1457. https://doi.org/10.1111/2041-210X.12788

Avcı, C., Sertel, E., & Kabadayı, M. E. (2022). Deep learning-based road extraction from historical maps. *IEEE Geoscience & Remote Sensing Letters*, *19*, 1–5. https://doi.org/10.1109/LGRS.2022.3204817

Bahgat, K., & Runfola, D. (2021). Toponym-assisted map georeferencing: Evaluating the use of toponyms for the digitization of map collections. *PLOS ONE*, *16*(11), 1–24. https://doi.org/10.1371/journal.pone.0260039

Budig, B., van Dijk, T. C., Feitsch, F., & Arteaga, M. G. (2016). Polygon consensus: Smart crowdsourcing for extracting building footprints from historical maps. In *Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (SIGSPACIAL '16)* (pp. 1–4). AMC.

Chen, Y., Carlinet, E., Chazalon, J., Mallet, C., Dumenieu, B., & Perret, J. (2021). Vectorization of historical maps using deep edge filtering and closed shape extraction. In J. Lladós, D. Lopresti, & S. Uchida (Eds.), *Document analysis and recognition – ICDAR 2021. ICDAR 2021. Lecture notes in computer science* (Vol. 12824, pp. 510–525). Cham: Springer. https://doi.org/10.1007/978-3-030-86337-1_34

Chiang, Y.-Y., Duan, W., Leyk, S., Uhl, J. H., & Knoblock, C. A. (2020). *Using historical maps in scientific studies: Applications, challenges, and best practices*. Springer Publishing Company, Incorporated.

Chiang, Y.-Y., & Knoblock, C. A. (2012). Generating named road vector data from raster maps. In N. Xiao, M. P. Kwan, M. F. Goodchild, & S. Shekhar (Eds.), *Geographic information science. GIScience 2012. Lecture notes in computer science* (Vol. 7478, pp. 57–71). Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-33024-7_5

Chiang, Y.-Y., & Knoblock, C. A. (2013). A general approach for extracting road vector data from raster maps. *International Journal on Document Analysis and Recognition (IJDAR)*, *16*(1), 55–81. https://doi.org/10.1007/s10032-011-0177-1

Chrysovalantis, G., & Tziokas, D. (2020). Building footprint extraction from historic maps utilizing automatic vectorisation methods in open source gis software. In I. Krisztina (Ed.), *Automatic vectorisation of historical maps* (pp. 9–17). International workshop organized by the ICA Commission on Cartographic Heritage into the Digital.

Ekim, B., Sertel, E., & Kabadayı, M. E. (2021). Automatic road extraction from historical maps using deep learning techniques: A regional case study of Turkey in a German World war *ii* map. *ISPRS International Journal of Geo-Information*, *10*(8), 1–15. https://doi.org/10.3390/ijgi10080492

Garcia-Molsosa, A., Orengo, H. A., Lawrence, D., Philip, G., Hopper, K., & Petrie, C. A. (2021). Potential of deep learning segmentation for the extraction of archaeological features from historical map series. *Archaeological Prospection*, *28*(2), 187–199. https://doi.org/10.1002/arp.1807

Gede, M., Árvai, V., Vassányi, G., Supka, Z., Szabó, E., Bordács, A., & Irás, K. (2020). Automatic vectorisation of old maps using qgis - tools, possibilities and challenges. In I. Krisztina (Ed.), *Automatic vectorisation of historical maps* (pp. 37–44). International workshop organized by the ICA Commission on Cartographic Heritage into the Digital.

Guo, Z., Shengoku, H., Wu, G., Chen, Q., Yuan, W., Shi, X., & Shibasaki, R. (2018). Semantic segmentation for urban planning maps based on u-net. In *IGARSS 2018–2018 IEEE International Geoscience And Remote Sensing Symposium*, Valencia, Spain (pp. 6187–6190). https://doi.org/10.1109/IGARSS.2018.8519049

Heitzler, M., & Hurni, L. (2020). Cartographic reconstruction of building footprints from historical maps: A study on the Swiss Siegfried map. *Transactions in GIS*, *24*(2), 442–461. https://doi.org/10.1111/tgis.12610

Jelena, I., Bojana, B., Aleksandar, R., & Dubravko, C. (2018). Deep learning for historical cadastral maps digitization: Overview, challenges and potential. In *26th International Conference on Computer Graphics, Visualization and Computer Vision 2018*, Václav Skala - UNION Agency) (pp. 42–47).

Jiao, C., Heitzler, M., & Hurni, L. (2020). Extracting wetlands from Swiss historical maps with convolutional neural networks. In *Automatic vectorisation of historical maps* (pp. 31–36). International workshop organized by the ICA Commission on Cartographic Heritage into the Digital.

Jiao, C., Heitzler, M., & Hurni, L. (2021). A survey of road feature extraction methods from raster maps. *Transactions in GIS*, *25*(6), 2734–2763. https://doi.org/10.1111/tgis.12812

Jiao, C., Heitzler, M., & Hurni, L. (2022a). A fast and effective deep learning approach for road extraction from historical maps by automatically generating training data with symbol reconstruction. *International Journal of Applied Earth Observation and Geoinformation*, *113*, 102980. https://doi.org/10.1016/j.jag.2022.102980

Jiao, C., Heitzler, M., & Hurni, L. (2022b). A novel data augmentation method to enhance the training dataset for road extraction from Swiss historical maps. *ISPRS Annals of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, *2*, 423–429. https://doi.org/10.5194/isprs-annals-V-2-2022-423-2022

Knechtel, J., Rottmann, P., Haunert, J.-H., & Dehbi, Y. (2024). Semantic floorplan segmenta- tion using self-constructing graph networks. *Automation in Construction*, *166*, 105649. https://doi.org/10.1016/j.autcon.2024.105649

Komac, B., & Gašperič, P. (2023). Cartographic time travel: Reflecting the past, defining the present, and challenging the future using old maps. *Acta Geographica Slovenica*, *63* (2), 9–21. https://doi.org/10.3986/AGS.13292

Laumer, D., Gűmgűmcű, H., Heitzler, M., & Hurni, L. (2020). A semi-automatic label digitization workflow for the Siegfried map. In I. Krisztina (Ed.), *Automatic vectorisation of historical maps* (pp. 55–62). International workshop organized by the ICA Commission on Cartographic Heritageinto the Digital.

le Riche, M. (2020). Identifying building footprints in historic map data using opencv and postgis. In I. Krisztina (Ed.), *Automatic vectorisation of historical maps* (pp. 18–30). International workshop organized by the ICA Commission on Cartographic Heritage into the Digital.

Li, Z. (2019). Generating historical maps from online maps. In F. Banaei-Kashani, T. Goce, H. G. Ralf, K. Lars, & N. Shawn (Eds.), *Proceedings of the 27th Acm Sigspatial International Conference on Advances in Geographic Information Systems* (pp. 610–611). Association for Computing Machinery.

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. In *Proceedings of the 26th Acm Sigkdd International Conference on Knowledge Discovery & Data Mining* (pp. 3290–3298). New York, NY, USA: Association for Computing Machinery.

Li, Z., Guan, R., Yu, Q., Chiang, Y.-Y., & Knoblock, C. A. (2021). Synthetic map generation to provide unlimited training data for historical map text detection. In L. Dalton, Y. Lexie, G. Song, M. Bruno, H. Yingjie, D. Xueqing, & N. Shawn (Eds.), *Proceedings of the 4th Acm Sigspatial International Workshop on Ai for Geographic Knowledge Discovery* (pp. 17–26). Association for Computing Machinery.

Liu, Q., Kampffmeyer, M., Jenssen, R., & Salberg, A.-B. (2020). Multi-view self-constructing graph convolutional networks with adaptive class weighting loss for semantic segmentation. In *Proceedings of the 2020 IEEE/CVF Conference On Computer Vision and Pattern Recognition Workshops (CVPRW 2020)* (pp. 199–205).

Liu, Q., Kampffmeyer, M., Jenssen, R., & Salberg, A.-B. (2021). Self-constructing graph neural networks to model long-range pixel dependencies for semantic segmentation of remote sensing images. *International Journal of Remote Sensing*, *42*(16), 6184–6208. https://doi.org/10.1080/01431161.2021.1936267

Luft, J. (2020). Automatic georeferencing of historical maps by geocoding. In I. Krisztina (Ed.), *Automatic vectorisation of historical maps* (pp. 75–88). Department of Cartography and Geoinformatics. ELTE Eötvös Loránd University.

Maxwell, A. E., Bester, M. S., Guillen, L. A., Ramezan, C. A., Carpinello, D. J., Fan, Y., & Pyron, J. L. (2020). Semantic segmentation deep learning for extracting surface mine extents from historic topographic maps. *Remote Sensing*, *12*(24), 4145. https://doi.org/10.3390/rs12244145

Mäyrä, J., Kivinen, S., Keski-Saari, S., Poikolainen, L., & Kumpula, T. (2023). Utilizing historical maps in identification of long-term land use and land cover changes. *AMBIO: A Journal of the Human Environment*, *52*(11), 1777–1792. https://doi.org/10.1007/s13280-023-01838-z

Milleville, K., Verstockt, S., & van de Weghe, N. (2020). Improving toponym recognition accuracy of historical topographic maps. In I. Krisztina (Ed.), *Automatic vectorisation of historical maps* (pp. 63–72). International workshop organized by the ICA Commission on Cartographic Heritage into the Digital.

Nobajas, A. (2020). Targeted crowdsourced vectorisation of historical cartography. In I. Krisztina (Ed.), *Automatic vectorisation of historical maps* (pp. 73). International workshop organized by the ICA Commission on Cartographic Heritage into the Digital.

Pack, C., & Kim, D. (2023). Weakly-supervised semantic segmentation on historical document images. *Proceedings of the 2023 International Conference on Research in Adaptive and Convergent Systems*, Association for Computing Machinery. https://doi.org/10.1145/3599957.3606219

Petitpierre, R. (2021). *Historical city maps semantic segmentation dataset*. Zenodo. https://doi.org/10.5281/zenodo.5513639

Petitpierre, R., Kaplan, F., & diLenardo, I. (2021). Generic semantic segmentation of historical maps. *Computational Humanities Research Conference*. Zenodo (pp. 228–248).

Piechl, A. (2020). (Semi-)automatic vector extraction of administrative borders from historical raster maps. In I. Krisztina (Ed.), *Automatic vectorisation of historical maps* (pp. 45–52). International workshop organized by the ICA Commission on Cartographic Heritage into the Digital.

Schlegel, I. (2019). Empirical study for a deployment of a methodology for improving the comparability between historical and current maps. *KN - Journal of Cartography and Geographic Information*, *69*(2), 121–130. https://doi.org/10.1007/s42489-019-00016-0

Schlegel, I. (2021). Automated extraction of labels from large-scale historical maps. *AGILE: GIScience Series*, *2*, 1–14. https://doi.org/10.5194/agile-giss-2-12-2021

Schlegel, I. (2023). A holistic workflow for semi-automated object extraction from large-scale historical maps. *KN - Journal of Cartography and Geographic Information*, *73*(1), 3–18. https://doi.org/10.1007/s42489-023-00131-z

Shbita, B., Knoblock, C. A., Duan, W., Chiang, Y.-Y., Uhl, J. H., & Leyk, S. (2023). Building spatio-temporal knowledge graphs from vectorized topographic historical maps. *Semantic Web*, *14*(3), 527–549. https://doi.org/10.3233/SW-222918

Sobotkova, A., Ross, S. A., Nassif-Haynes, C., & Ballsun-Stanton, B. (2023). Creating large, high-quality geospatial datasets from historical maps using novice volunteers. *Applied Geography*, *155*, 102967. https://doi.org/10.1016/j.apgeog.2023.102967

Uhl, J. H., Leyk, S., Chiang, Y.-Y., & Knoblock, C. A. (2022). Towards the automated large- scale reconstruction of past road networks from historical maps. *Computers, Environment and Urban Systems*, *94*, 101794. https://doi.org/10.1016/j.compenvurbsys.2022.101794

Uhl, J. H., Leyk, S., Li, Z., Duan, W., Shbita, B., Chiang, Y.-Y., & Knoblock, C. A. (2021). Combining remote-sensing-derived data and historical maps for long-term back-casting of urban extents. *Remote Sensing*, *13*(18), 1–25. https://doi.org/10.3390/rs13183672

Weinman, J. (2013). Toponym recognition in historical maps by gazetteer alignment. In *2013 12th International Conference on Document Analysis and Recognition* (pp. 1044–1048). IEEE. https://doi.org/10.1109/ICDAR.2013.209

Wu, S., Chen, Y., Schindler, K., & Hurni, L. (2023). Cross-attention spatio-temporal context transformer for semantic segmentation of historical maps. In *Proceedings of the 31st Acm International Conference on Advances in Geographic Information Systems*. Association for Computing Machinery.

Wu, S., Heitzler, M., & Hurni, L. (2022). Leveraging uncertainty estimation and spatial pyramid pooling for extracting hydrological features from scanned historical topographic maps. *GIScience & Remote Sensing*, *59*(1), 200–214. https://doi.org/10.1080/15481603.2021.2023840