



Extracting road features from historical  
maps by incorporating cartographic  
method knowledge into deep learning



Chenjing Jiao  
ETH ZÜRICH

DISS. ETH NO. 30029



DISS. ETH NO. 30029

Extracting road features from historical maps by incorporating  
cartographic method knowledge into deep learning

A thesis submitted to attain the degree of  
DOCTOR OF SCIENCES of ETH ZURICH  
(Dr. sc. ETH Zürich)

presented by  
Chenjing Jiao

M.Sc. in Photogrammetry and Remote Sensing, Wuhan University, China

accepted on the recommendation of  
Prof. Dr. Lorenz Hurni  
Prof. Dr. Jochen Schiewe  
Dr. Julien Perret

2024



## Acknowledgments

I am profoundly grateful to my doctoral supervisor, Prof. Dr. Lorenz Hurni, for his unwavering mentorship and invaluable insights throughout the research process. His expertise and dedication have been instrumental.

I extend my sincere appreciation to the members of the doctoral examination committee, Prof. Dr. Konrad Schindler, Dr. Julien Perret, and Prof. Dr. Jochen Schiewe, for their constructive feedback and expert evaluation of my work.

I am indebted to my fellow researchers and colleagues at the Institute of Cartography and Geoinformation, ETH Zürich for their collaboration and stimulating discussions. Their collective insights and encouragement have been invaluable. I want to express special thanks to Dr. Magnus Heitzler for his mentorship, which has contributed to my research, teaching, and scientific writing abilities, etc. Over the past years, we have engaged in numerous discussions, brainstormed ideas, and collaborated on papers and projects.

My heartfelt gratitude goes to my parents, my grandmother and my husband, for their unwavering support and understanding during the challenging phases of this journey. To my friends, I thank you for being a constant source of solace, even though some of you are located 9,000 kilometers away.

I also acknowledge the contributions of previous researchers in this field, whose work has laid the foundation and served as inspiration for my own research.

Thank you all!

04. Nov. 2023



## Abstract

Road network data are crucial to various scientific domains, including urban planning and transportation. They find wide-ranging applications owing to the valuable geo-spatial and network information they encompass. Notably, road network data that span extended time periods contain additional temporal information, making them indispensable for multi-temporal analyses such as investigations into road network evolutions and urban growth processes. Moreover, long-term road network data play a pivotal role in the creation of 3D historical cityscapes. Nevertheless, how to obtain multi-temporal road network data remains a challenge due to limited data sources. Historical documentations of transportation infrastructures record constructions and alterations of roads, but lacks precise spatial information.

Historical maps can serve as an invaluable source of long-term road network data, which are often temporally successive and spatially accurate, particularly from the 19<sup>th</sup> century onwards. Although abundant historical maps have been georeferenced, digitalized and made accessible online, valuable information remains hidden in them due to challenges in processing these map images, because they are usually subject to unsatisfactory quality. Among all feature extraction tasks, road extraction can prove to be particularly challenging due to its large geographical extent and interactions with other features.

Numerous methods have been proposed and developed for road extraction from overhead imagery and raster maps. Particularly, machine/deep learning techniques play an increasingly pivotal role in various image processing tasks, including feature extraction. Nonetheless, limited studies have been conducted in utilizing deep learning for road extraction from historical maps. Furthermore, to investigate the existing road extraction methods, a comprehensive literature survey is conducted to review road extraction methods from raster maps of the last 35 years as well as recent methods from overhead imagery. However, it is found that in existing studies, no cartographic method knowledge is incorporated in deep learning-based road extraction from historical maps, despite their inherent cartographic nature.

To bridge this research gap as well as to produce high-quality long-term road data, this thesis is aiming at developing a comprehensive framework that incorporates cartographic method knowledge into deep learning-based road extraction from historical maps. The produced road datasets include raster road segmentations, vectorized road line and road classifications. Specifically, first a novel data augmentation method is proposed and implemented, which integrates cartographic symbolization knowledge by transforming target features only (roads in this case) instead of the whole image patch. Deep learning models trained with this novel method outperform those without data augmentation. Moreover, a method for automatically creating training data is developed through symbol reconstruction, which employs cartographic colour use and symbolization knowledge. Models either trained with these automatically generated training data or with mixed automatically and manually labeled data produce promising road segmentation results. These two methods complement or replace the effort of purely manually labeling training data. Additionally, a method for classifying and vectorizing the road segmentations is developed through symbol painting, which incorporates symbolization knowledge into deep learning. Road segmentations and vectorization results verify the effectiveness of integrating cartographic method knowledge into deep learning. Although these three methodologies are validated on road extraction from Swiss Siegfried map, they can be generalized to other feature extraction tasks and to other map series. The produced road segmentations, road lines and classifications can be used for various analyses such as road network evolutions, urban sprawl and sustainable transportation.



## Zusammenfassung

Strassennetzwerk dokumentationen sind entscheidend für verschiedene wissenschaftliche Bereiche, einschliesslich der Stadtplanung und den Verkehrswissenschaften. Sie finden vielfältige Anwendungen aufgrund der wertvollen geografischen und netzwerkbezogenen Informationen, die sie enthalten. Insbesondere Strassennetzwerksdaten, die sich über lange Zeiträume erstrecken, enthalten zusätzliche zeitliche Informationen und sind daher für multi-temporale Analysen wie Untersuchungen zur Entwicklung von Strassennetzwerken und städtischen Wachstumsprozessen unerlässlich. Darüber hinaus spielen langfristige Strassennetzwerksdaten eine entscheidende Rolle bei der Erstellung von historischen 3D-Stadtmodellen. Dennoch bleibt die Beschaffung von multi-temporalen Strassennetzwerk dokumentationen aufgrund begrenzter Datenquellen eine Herausforderung. Historische Dokumentationen von Verkehrsinfrastrukturen erfassen den Bau und die Veränderungen von Straßen, verfügen jedoch nicht über genaue räumliche Informationen.

Historische Karten können als unschätzbar wertvolle Quellen für langfristige Strassennetzwerk dokumentationen dienen, die oft zeitlich aufeinanderfolgend und räumlich genau sind, insbesondere ab dem 19. Jahrhundert. Obwohl zahlreiche historische Karten georeferenziert, digitalisiert und online zugänglich gemacht wurden, bleiben wertvolle Informationen in ihnen aufgrund von Herausforderungen bei der Verarbeitung dieser Kartenbilder verborgen, da sie in der Regel von unbefriedigender Qualität sind. Unter allen Aufgaben zur Merkmalsextraktion kann die Strassenextraktion aufgrund ihres umfangreichen Charakters und ihrer Wechselwirkungen mit anderen Merkmalen besonders anspruchsvoll sein.

Es wurden zahlreiche Methoden für die Strassenextraktion aus Luftbildern und Rasterkarten vorgeschlagen und entwickelt. Insbesondere spielen Machine/Deep Learning-Techniken eine zunehmend zentrale Rolle in verschiedenen Bildverarbeitungsaufgaben, einschliesslich der Merkmalsextraktion. Dennoch wurden nur begrenzte Studien zur Nutzung von Deep Learning für die Strassenauswahl aus historischen Karten durchgeführt. Darüber hinaus wurde eine umfassende Literaturrecherche durchgeführt, um die vorhandenen Strassenauswahlmethoden zu untersuchen. Diese umfasste die Bewertung von Strassenauswahlmethoden aus Rasterkarten der letzten 35 Jahre sowie neueren Methoden aus Luftbildern. Es wurde jedoch festgestellt, dass in den vorhandenen Studien kein kartografisches Methodenwissen in die auf Deep Learning basierende Strassenauswahl aus historischen Karten integriert ist, obwohl diese aufgrund ihrer kartografischen Natur.

Um die bestehende Lücke in der Forschung zu schliessen und hochwertige, langfristige Strassendaten zu generieren, zielt diese Arbeit darauf ab, ein umfassendes Rahmenwerk zu entwickeln, das kartografisches Methodenwissen in die auf Deep Learning basierende Extraktion aus historischen Karten integriert. Die erstellten Strassendatensätze umfassen Raster-Strassensegmentierungen, vektorisierte Straßenlinien und Straßenklassifikationen. Namentlich wird zuerst eine neuartige Methode zur Datenanreicherung vorgeschlagen und umgesetzt, die kartografisches Symbolwissen integriert, indem nur Zielmerkmale (in diesem Fall Straßen) anstelle des gesamten Bildausschnitts transformiert werden. Mit dieser neuartigen Methode trainierte Deep Learning-Modelle übertreffen diejenigen, die ohne Datenanreicherung trainiert wurden. Darüber hinaus wird eine Methode zur automatischen Erstellung von Trainingsdaten durch Symbolrekonstruktion entwickelt, die auf der Anwendung von kartografischem Wissen zu Farbgebung und Symbolisierung basiert. Modelle, die entweder mit diesen automatisch generierten Trainingsdaten oder mit gemischten automatisch und manuell gelabelten Daten trainiert sind, erzielen vielversprechende Ergebnisse bei der Strassensegmentierung. Diese beiden Methoden können das manuelle Erzeugen von Trainingsdaten ergänzen oder sogar ersetzen. Zusätzlich wird eine Methode zur Klassifizierung und Vektorisierung der Strassensegmentierungen durch «Symbolzeichnen» entwickelt, die kartografisches Symbolwissen in das Deep Learning integriert. Die Strassensegmentierungen und die Ergebnisse der Vektorisierung bestätigen die Effektivität der Integration von kartografischem Methodenwissen in das Deep Learning.

Obwohl diese drei Methoden in der Strassenextraktion aus der Schweizer Siegfried-Karte validiert sind, können sie auf andere Merkmalsextraktionsaufgaben und andere Kartenserien angewendet werden. Die erstellten Strassensegmentierungen, Strassenlinien und Klassifikationen können für verschiedene Analysezwecke, wie die Untersuchung der Entwicklung von Strassennetzen, städtisches Wachstum und nachhaltigen Verkehr, genutzt werden.

## Table of Contents

Acknowledgments .....	i
Abstract .....	iii
Zusammenfassung .....	v
Table of Contents .....	vii
Publications .....	xi
1. Introduction .....	1
1.1. Motivation and problem statement .....	2
1.2. Research questions and methodology .....	3
1.3. Relevance for science and society .....	4
1.4. Organization of the thesis .....	5
2. Background .....	7
2.1. A brief introduction to historical maps .....	7
2.1.1. Historical Swiss maps .....	7
2.1.2. Selected foreign historical maps .....	10
2.2. Digitalization, georeferencing and access of historical maps .....	10
2.3. Quality issues in historical maps .....	12
2.3.1. Quality issues .....	13
2.3.1.1. Production-relevant issues .....	13
2.3.1.2. Preservation-relevant issues .....	15
2.3.1.3. Transformation-relevant issues .....	16
2.3.2. Assessing and visualizing inaccuracies .....	18
2.4. Segmentation and vectorization of historical maps .....	20
2.4.1. Road and railway extraction .....	20
2.4.2. Label extraction .....	21
2.4.3. Contour line extraction .....	21
2.4.4. Human settlement extraction .....	21
2.4.5. Hydrological feature extraction .....	22
2.5. Applications of historical maps .....	22
2.6. Techniques for historical map image processing .....	23
2.6.1. Rectification and resampling .....	23
2.6.2. Colour image segmentation .....	23
2.6.2.1. Histogram thresholding .....	24
2.6.2.2. Colour space clustering .....	24
2.6.3. Morphological operations .....	25
2.6.4. Template matching .....	26
2.6.5. Optical character recognition .....	26

2.6.6. Machine learning .....	27
2.6.7. Raster-to-vector conversion .....	29
3. A survey of road feature extraction methods from raster maps .....	31
Abstract .....	31
3.1 . Introduction .....	31
3.2 . Road feature extraction from raster maps .....	33
3.2.1 . Categorization.....	33
3.2.2 . Line extraction-based road extraction .....	38
3.2.2.1 . Line tracing.....	38
3.2.2.2 . Parallel characteristic extraction .....	41
3.2.2.3 . Hough transform.....	41
3.2.3 . Image filtering-based road extraction .....	42
3.2.3.1 . Conventional filters .....	42
3.2.3.2 . Morphological operations .....	42
3.2.3.3 . Convolutional neural network.....	44
3.2.4 . CIS-based road extraction.....	45
3.2.4.1 . Histogram technique .....	45
3.2.4.2 . K-means.....	46
3.2.4.3 . Other clustering techniques.....	47
3.2.4.4 . Support vector machine .....	48
3.3 . Road feature extraction from overhead imagery .....	49
3.3.1. Machine learning architectures .....	49
3.3.2. Alternative loss functions .....	50
3.3.3. Data fusion .....	50
3.3.4. Shape descriptors .....	51
3.3.5. Evaluation metrics .....	51
3.4 . Discussion .....	52
3.4.1. An analysis of the development and progress of the road extraction methods of raster maps .....	52
3.4.2. Current limitations and technology trends .....	55
3.5 . Conclusions .....	55
Appendix .....	56
References .....	61
4. A novel data augmentation method to enhance the training dataset for road extraction from Swiss historical maps .....	69
Abstract .....	69
4.1. Introduction .....	69
4.2. Data and method.....	70
4.2.1 . Data .....	70

4.2.2 . Sampling strategy .....	72
4.2.3 . A novel data augmentation method .....	73
4.2.4 . Road extraction with U-Net.....	74
4.3. Experiment .....	75
4.3.1 . Training scenarios.....	75
4.3.2 . Post-processing.....	75
4.3.3 . Evaluation.....	77
4.4. Conclusion.....	78
Acknowledgements .....	78
References .....	79
5. A fast and effective deep learning approach for road extraction from historical maps by automatically generating training data with symbol reconstruction.....	81
Abstract .....	81
5.1 . Introduction .....	81
5.2 . Related work .....	83
5.2.1 . Road extraction using deep CNNs .....	83
5.2.2 . Automatic generation of training data.....	83
5.3 . Data .....	84
5.3.1 . Siegfried map and the symbols in the black layer .....	84
5.3.2 . The available road training data.....	85
5.3.3 . The contemporary vector data .....	86
5.4 . Methodology .....	87
5.4.1. Symbol reconstruction to automatically generate training data.....	87
5.4.2. Road extraction with U-net.....	91
5.4.2.1 . Network architecture.....	91
5.4.2.2 . Training .....	92
5.4.3. Extracting the black layer from Siegfried map .....	94
5.5 . Experiment .....	95
5.5.1 . Road extraction.....	95
5.5.2 . Evaluation.....	96
5.6 . Discussion and conclusion.....	100
Acknowledgement.....	102
Appendix .....	102
References .....	103
6. A novel framework for road vectorization and classification from historical maps based on deep learning and symbol painting .....	107
Abstract .....	107
6.1 . Introduction .....	107
6.2 . Data .....	109

6.3 . Methodology .....	111
6.3.1 . Road segmentation .....	111
6.3.2 . Road skeletonization.....	113
6.3.3 . Road vectorization and classification.....	114
6.4 . Experiment .....	118
6.4.1 . Painting process.....	119
6.4.2 . Visual assessment .....	120
6.4.3 . Quantitative assessment.....	123
6.4.4 . Sensitivity analysis .....	124
6.5 . Discussion .....	126
6.6 . Conclusion.....	127
References .....	127
7 . Conclusions .....	133
7.1 . Synthesis and Contributions .....	133
7.2 . Outlook for Future Research .....	134
7.2.1 . Method-oriented directions .....	134
7.2.2 . Application-oriented directions .....	136
Reference .....	137
List of Figures .....	149
List of Tables.....	151

## Publications

### Contributions included in this dissertation

**Jiao, C.**, Heitzler, M. and Hurni, L., 2021. A survey of road feature extraction methods from raster maps. *Transactions in GIS*, 25(6), pp.2734-2763.

**Jiao, C.**, Heitzler, M. and Hurni, L., 2022. A novel data augmentation method to enhance the training dataset for road extraction from Swiss historical maps. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2, pp.423-429.

**Jiao, C.**, Heitzler, M. and Hurni, L., 2022. 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, p.102980.

**Jiao, C.**, Heitzler, M. and Hurni, L., 2024. A novel framework for road vectorization and classification from historical maps based on deep learning and symbol painting. *Computers, Environment and Urban Systems*, 108, p.102060.

### Contributions not included in this dissertation

Zhao, P., Jia, T., Qin, K., Shan, J. and **Jiao, C.**, 2015. Statistical analysis on the evolution of OpenStreetMap road networks in Beijing. *Physica A: Statistical Mechanics and its Applications*, 420, pp.59-72.

**Jiao, C.**, Heitzler, M. and Hurni, L., 2020. Extracting Wetlands from Swiss Historical Maps with Convolutional Neural Networks. In *Automatic Vectorisation of Historical Maps. International workshop organized by the ICA Commission on Cartographic Heritage into the Digital 13 March, 2020 Budapest. Proceedings* (pp. 33-38). Department of Cartography and Geoinformatics, ELTE Eötvös Loránd University.

**Jiao, C.**, Heitzler, M. and Hurni, L., 2021. Utilizing convolutional neural networks to extract road features from Swiss historical maps. *Abstracts of the ICA*, 3, pp.1-2.

**Jiao, C.** and Hurni, L., 2021, December. A survey on the evolution of the Chinese cadastral system. In *Proceedings of the ICA* (Vol. 4, p. 47). Göttingen, Germany: Copernicus Publications.

Räth, Y.M., Grêt-Regamey, A., **Jiao, C.**, Wu, S. and van Strien, M.J., 2023. Settlement relationships and their morphological homogeneity across time and scale. *Scientific Reports*, 13(1), p.11248.

Xia, X., **Jiao, C.**, and Hurni, L., 2023. Contrastive Pretraining for Railway Detection: Unveiling Historical Maps with Transformers. In *6<sup>th</sup> ACM SIGSPATIAL International Workshop on AI for Geographic Knowledge Discovery (GeoAI '23)*, November 13, 2023, Hamburg, Germany. ACM, New York, NY, USA.



## 1. Introduction

Road network data are of paramount importance in various scientific disciplines due to their crucial role in enhancing research, analysis, and decision-making processes. The geo-spatial information and complex-network knowledge contained in road network data lay indispensable basis for studies in urban planning, transportation engineering, human geography, environmental science, and many more. (Chiang *et al.*, 2005; Strano *et al.*, 2017). For example, road networks, as integral components of urban planning and transportation systems, play a vital role in shaping the functionality, efficiency, and sustainability of cities. Particularly, long-term historical road network data are fundamental to analyses of the evolution of road networks, investigations into the process of urban sprawl and implementations of multi-temporal geo-spatial data fusion, as road networks exist across various geo-spatial data sources (Erath *et al.*, 2009; Axhausen *et al.*, 2011; Strano *et al.*, 2012; Masucci *et al.*, 2014; Zhao *et al.*, 2015; Strano *et al.*, 2017; Saeedimoghaddam and Stepinski, 2020). Moreover, historical road network data from the past are used to reconstruct historical cityscapes for educational, entertainment, and research purposes<sup>1</sup> (Jiao *et al.*, 2021). All of these research and applications necessitate road network data that cover a long time-span.

Historical maps are essential and sometimes the only data source of historical road network data. For example, Peutinger's Tabula, or Tabula Peutingeriana<sup>2</sup>, is an ancient Roman map depicting the road network of the Roman Empire. Roads are represented with orange-coloured single lines connecting settlements (Albu *et al.*, 2014). Tabula Rogeriana, or The Book of Roger<sup>3</sup>, is an atlas of the 12<sup>th</sup> century completed by the Arabic geographer Muhammad al-Idrisi, in which roads are shown as green lines (Haque, 2018). Later, starting in the 19<sup>th</sup> century, with the advent of advanced surveying and mapping technologies, topographical maps with a higher spatial and semantic precision were developed. They were usually regularly updated, thus providing a continuous record of road data over time. For instance, in historical United States Geological Survey (USGS) maps<sup>4</sup>, the Topographic Atlas of Switzerland or the Siegfried Map<sup>5</sup> and Swiss Old National map, road networks are clearly and precisely depicted, with different symbols representing different road classes.

However, paper-based historical maps usually cannot be directly used for analyses. To take advantages of them in the contemporary digital age, they first have to be digitalized, namely they must be scanned into raster format that is represented as a grid of pixels. These digitalized historical map images can then easily be stored, visualized and distributed. Subsequently, the historical map images need to be georeferenced, i.e., to determine the real-world location of the scanned map (content) in respect to a reference coordinate system. Georeferencing makes it possible to superimpose the map with other geo-spatial data. Furthermore, feature extraction is concerned with deriving geographical features from the georeferenced historical maps. Features can be extracted into raster and later converted into vector format. This facilitates the analyses and comparison of geographical features across time. Particularly, spatio-temporal analyses and geoprocessing operations (e.g., buffering, clipping) can be used on the extracted features in a Geographical Information System (GIS) in order to investigate the evolution of the features and the interactions between features (e.g., roads and settlements) (Räth *et al.*, 2023). The extracted features from historical maps will benefit researchers and analysts from, but not limited to, domains including urban planning, transportation, ecology, and toponymy (Jiao *et al.*, 2020).

<sup>1</sup> <https://blog.research.google/2020/10/recreating-historical-streetscapes.html>

<sup>2</sup> [https://en.wikipedia.org/wiki/Tabula\\_Peutingeriana](https://en.wikipedia.org/wiki/Tabula_Peutingeriana)

<sup>3</sup> [https://en.wikipedia.org/wiki/Tabula\\_Rogeriana](https://en.wikipedia.org/wiki/Tabula_Rogeriana)

<sup>4</sup> [https://en.wikipedia.org/wiki/United\\_States\\_Geological\\_Survey](https://en.wikipedia.org/wiki/United_States_Geological_Survey)

<sup>5</sup> [https://en.wikipedia.org/wiki/Topographic\\_Atlas\\_of\\_Switzerland](https://en.wikipedia.org/wiki/Topographic_Atlas_of_Switzerland)

### 1.1. Motivation and problem statement

Various natural and anthropogenic geographical features are depicted in historical maps, such as lakes, rivers, buildings, road networks. A range of methods have been proposed and developed for feature extraction from historical maps. These methods can roughly be categorized into conventional methods and machine learning methods. Conventional methods include colour image segmentation (CIS), morphological operations, line tracing, template matching, etc. For example, interactive or automatic line-tracing methods can be used for extracting linear features like contour lines and roads. Machine/deep learning methods offer an effective way to extract features from historical maps due to their excellent performance in image processing. They have been used for extracting building footprints (Heitzler and Hurni, 2020), hydrological features (Jiao *et al.*, 2020; Wu *et al.*, 2022), and roads (Chiang *et al.*, 2020a; Duan *et al.*, 2020; Saeedimoghaddam and Stepinski, 2020; Avci *et al.*, 2022).

Conventional methods are usually relatively easy to implement, as there are already various conventional computer vision techniques available that have been developed over the years. Moreover, it is usually easier to understand how conventional methods work compared to machine/deep learning methods. However, these conventional methods often rely heavily on manually fine-tuned parameters (Leyk, and Boesch, 2009), and usually a set of conventional techniques needs to be combined to accomplish a feature extraction task from historical maps. Additionally, these conventional methods usually have limited generalizability (Chiang and Knoblock, 2014; Jiao *et al.*, 2020), and may require manual intervention (Chiang and Knoblock, 2009a). In contrast, machine/deep learning methods present higher performance and generalizability in feature extraction when sufficient high-quality training data are available. Different from manually designing appropriate filters or descriptors to extract features in conventional methods, machine/deep learning automatically learns diverse and representative features from training data, leading to more effective feature extraction (Long *et al.*, 2015; Ronneberger *et al.*, 2015). However, due to the data-driven nature of deep learning, its major disadvantage lies in the need for extensive training data. Producing such data is typically time-consuming and labour-intensive. Additionally, deep learning models are often described as “black boxes”, thus understanding how they make predictions or decisions is still challenging.

Both conventional methods and machine/deep learning techniques have been used for feature extraction from raster maps (e.g., scanned historical maps, contemporary maps in raster format) and overhead imagery. Although overhead imagery is different from historical maps, methods applied to it can be tailored to historical maps due to their similar geo-spatial nature and data format. The focus of this dissertation is laid on road networks. Owing to the importance and wide use of road network data in various scientific disciplines, this dissertation aims at contributing to the body of knowledge of feature extraction from historical maps by focusing specifically on the challenges and potential solutions for the extraction of road networks. Thus, it's necessary to make a thorough survey of these conventional and machine/deep learning methods for road extraction from raster maps. Moreover, recent advances of machine/deep learning methods applied to overhead imagery also need to be reviewed, as they will possibly offer valuable insights for road extraction from raster maps. The survey will be helpful to investigate where conventional methods are sufficient and where deep learning methods are useful in road extraction.

As cartographic products, historical maps contain a wealth of cartographic method knowledge (e.g., symbolization rules, generalization processes), which sets them apart from other geo-spatial data such as overhead imagery. Although deep learning has been attracting notable attention in feature extraction from scanned historical maps (Chiang *et al.*, 2020a), little effort has been put on investigating how cartographic method knowledge can be used to improve feature extraction results using deep learning methods.

## 1.2. Research questions and methodology

To develop and implement road extraction methods from historical maps, it makes sense to review existing methods and to learn state-of-the-art techniques. In addition, a thorough understanding of existing methods makes it possible to point out their limitations and identify possible future directions for development. Although techniques applied in road extraction evolve through time driven by the continuous advancement of computer science technologies, to the best of our knowledge, there has been no comprehensive literature review of road extraction methods from raster maps. More attention, however, has been put on reviewing those methods applied to overhead imagery. Thus, the 1<sup>st</sup> research question addressed in this dissertation is

Research Question 1: Which methods exist to extract road features from raster maps and overhead imagery?

Here we broaden the scope of data source to raster maps and overhead imagery, as only a limited amount of methods have been developed specifically for road extraction from historical maps. Apart from this, methods applied to raster maps beyond historical maps as well as recent advances in overhead imagery will possibly benefit the development of innovative methods for historical maps.

Although deep learning has been widely applied for feature extraction from images (e.g., overhead imagery), little attention has been paid to developing deep learning-based methods for road extraction from historical maps. There exist some studies that have focused on extracting natural land objects (e.g., forests, wetlands) from historical maps (Leyk and Boesch, 2009; Loran *et al.*, 2018). However, methods for anthropogenic features, especially road infrastructures, have not been investigated, despite of their wide use across scientific domains and applications. Road extraction from historical maps can be challenging. This difficulty stems from the resemblance of road symbols to other feature symbols like isolines and streams, the extensive length of roads, the intersection and adjacency of roads with other features of the same colour (e.g., buildings), the marginal differences between symbols representing different road classes (e.g., main roads, country roads), the unsatisfactory quality of scanned historical maps that have been printed decades or even centuries ago (e.g., bleaching, blurring and colour aliasing) (Jiao *et al.*, 2021).

A major gap that has been identified through this literature review is that there exists no study that utilizes cartographic method knowledge to enhance deep learning for feature extraction from historical maps. However, this aspect is deemed to be crucial when dealing with historical maps, and thus motivates the overarching angle of this dissertation. To fill this gap, Research Questions 2-4 are proposed.

The first step of exploiting supervised deep learning is to create sufficient amounts of training data. However, it is usually time-consuming and tedious to label training data, the amount and quality of which have great impact on the training performance. Specifically for roads, the complexity of road networks as well as the demand for sufficient training data for each road class makes the issue even more challenging. An effective solution to boost a training dataset for deep learning is data augmentation. Thus, the 2<sup>nd</sup> research question tackled in this dissertation is as follows:

Research Question 2: How can cartographic method knowledge be employed to augment training data for deep-learning-based road extraction from historical maps?

This research question addresses the scenario in which there are some road training data available, how data augmentation can be used to boost the training dataset, thereby improving road extraction results. Conventional data augmentation methods transform each training tile as a whole (e.g., rotation, flipping), in which no cartographic method knowledge (e.g., label, symbol) is employed. However, some features or symbols on maps will never occur in practice when they are rotated or flipped (e.g., numbers, labels).

Thus, a novel data augmentation method that will leverage the cartographic method knowledge as well as further improve the diversity of training samples will be implemented.

For scenarios in which no training data is available, it makes sense to investigate how to automatically create training data, i.e., to create synthetic training data. Thus, the 3<sup>rd</sup> question tackled is

**Research Question 3:** How can cartographic method knowledge be used to automatically create training data for deep learning-based road extraction from historical maps?

Deep learning methods automatically learn different levels of feature representations. On (historical) maps, different features are represented with distinct symbols. Thus, symbols can serve as a valuable tool for automatically generating training data. Moreover, these automatically-generated data can be combined with manually labeled road data for training. Models will be trained with the automatically-generated data only, with the manually-labeled data only, and with the combination of the two datasets, respectively. Road extraction results of these models will be analyzed and compared. The comparison will reveal how cartographic method knowledge helps enhance the capabilities of deep learning.

Following Research Question 3, in which pixel-wise road predictions are made from historical maps, a practical issue is how to vectorize the raster results and how to classify the vectorized road lines. On one hand, conventional vectorization methods usually apply morphological operations to first get the skeletons of road pixels and then utilize junction/network analysis to track the skeletons in order to obtain lines. These methods have high time and space complexity (Chen *et al.*, 2015). Additionally, usually these conventional methods produce saw-toothed representations of the vectorized road centerlines. On the other hand, although supervised deep learning achieves promising results in image classification and semantic segmentation, one major issue is that it requires large amounts of training data. The issue is even harsher for road classification, as enough training data are required for each road class. Thus, the 4<sup>th</sup> research question addressed in this dissertation is

**Research Question 4:** How can cartographic method knowledge help improve deep learning-based-road vectorization and classification results?

To address this research question, a workflow is introduced that incorporates cartographic road symbols to enhance road classification. Specifically, the training data available are road vector lines, which are then rasterized for training. The vector lines are not classified into specific road classes. Predictions from deep CNNs are in raster format, which have to be vectorized and classified. The crucial point of vectorization lies in finding turning points on road curves and approximating curves with sequentially connected segments. Classification will be implemented by employing cartographic symbolization knowledge, as different road classes are distinguished with different symbols.

Although methodologies stated above are validated on road extraction from Swiss historical maps, they can be generalized and applied to other feature extraction tasks for other historical map series.

### 1.3. Relevance for science and society

The vectorized road network data allow for interactive operation, analysis and management in GIS, thereby contributing to multiple domains in science and society. One important use of long-term road data is to analyze the evolution of road networks. Strano *et al.* (2012) explore the underlying mechanisms and principles governing the growth and evolution of road networks through quantitative spatio-temporal analyses of road data covering nearly 200 years in Milan, Italy. The analyses of road network evolution reveal urban development patterns. For example, Masucci *et al.* (2015) investigate urban growth patterns with road network data that cover more than 200 years. They present universal properties of cities and demonstrate how the natural boundaries of cities can be defined mathematically. Furthermore, the analyses of evolution of road networks reveal valuable insights into the socio-economic

changes in a region, such as changes in trade routes, economic centers and population shifts, from which historians, archaeologists and social scientists will benefit (Chaudhuri and Clarke, 2015). Costa *et al.* (2021) study how investments in roads and rail transits influences the vertical and horizontal urban development by taking São Paulo, Brazil as a case study. It is reported that the construction of avenues and arterial roads that cross the urbanized area and connect suburban neighborhoods has led to urban expansion.

The rapid advances in technology, such as the Internet of Things (IoT) (Atzori *et al.*, 2010) and Location-Based Services (LBS) (Raper *et al.*, 2007), have made it possible to access a wide range of geo-spatial information. Road network data from historical maps provide new and great opportunities to facilitate multi-temporal geo-spatial data fusion for addressing various challenges in urban, social, and environmental sustainability. For instance, Shirvani *et al.* (2020) investigate the spatial variations of Hyrcanian forest loss, fragmentation, and degradation due to the expansion of rural, logging, and mine roads. Various geo-spatial data from aerial photos, topographic maps, Landsat data and Google Earth images were collected and fused to generate a precise road network between 1966 and 2016 in northeast Iran. Historical maps can be a perfect supplement for generating road networks, especially long-term time series. In addition, historical maps have a potential to provide more attribute information of road networks, such as road width and road class.

Long-term road data extracted from historical maps are valuable to society, especially for education and cartographic heritage preservation. The road networks and their development can be used to illustrate major historical events and changes, such as migrations, the growth of towns and cities, or the impact of significant historical events like wars or industrialization. In addition, students can learn how terrain and natural resources influence road construction, how road networks connect resources, towns and cities, and how they facilitate trade and communication (Powers *et al.*, 2014).

Data extraction contributes to digital documentation and cultural preservation of paper-based historical maps, which belong to the cartographic heritage (Velikić *et al.*, 2022). Historical maps are subject to physical degradation over time, and requirements for preserving them are strict (e.g., maintaining specified temperature and humidity levels) (Baty *et al.*, 2010). Additionally, paper maps sometimes cannot be well preserved when data-holding institutes are reorganized or downsized, and when map holders themselves retire or pass away. This issue is acute in developing and transitional countries (Panagos *et al.*, 2011). However, keeping data in digital format extracted from historical maps is far easier. Moreover, these data can usually be easily accessed by the public as well as future generations. This accessibility fosters a deeper appreciation of the historical cartographic heritage.

#### 1.4. Organization of the thesis

The structure of the dissertation is outlined across seven chapters, each of which is described below.

In Chapter 2, we elaborate on the background knowledge of historical maps. First, a brief overview of historical maps is given. Second, we expound on how historical maps can be digitalized, georeferenced and accessed. Moreover, historical maps are susceptible to various quality issues due to limitations of the surveying, mapping and printing techniques applied in the past. Thus, we list and summarise these quality issues, and try to find possible reasons of these issues. Then, considering that the application of historical maps necessitates the segmentation and vectorization of geographical features, methods of segmentation and vectorization from historical maps are briefly reviewed. Furthermore, various application domains of historical maps are reviewed, with a particular emphasis on the use of road data extracted from historical maps. Last but not least, we summarise techniques employed in historical map image processing.

The goal of Chapter 3 is to provide a thorough understanding of the existing road extraction methods from raster maps, which is presented in the journal paper by Jiao *et al.* (2021). As laid out in this paper, “The methods are categorized from different perspectives, including the classes of employed techniques (e.g., line extraction, line tracing, Hough transform), the amount of required user intervention (e.g., interactive, automatic), the required data (e.g., scanned maps, contemporary vector data) and the produced results (e.g., raster-based predictions, vector-based results, attributes)”. Moreover, road extraction methods from overhead imagery published in recent years, which will possibly contribute to road extraction from raster maps, are reviewed. Additionally, various evaluation metrics for road extraction and vectorization results are described. More importantly, the evolution of these methods over the past 35 years is analysed, and their limitations, as well as possible future development directions, are discussed. This chapter answers the Research Question 1.

Chapter 4 focuses on examining the feasibility of deep learning in extracting roads from historical maps as well as dealing with the obstacles of manually labelling large amounts of training data, which is time-consuming and laborious. To achieve the goals, deep convolutional neural networks (CNNs) are developed and implemented, and a novel data augmentation method is proposed to enhance the training dataset. Compared with the existing data augmentation techniques that create training data examples by transforming each training sample/patch as a whole (e.g., rotation, flipping), we investigate how to further improve the diversity of training samples by varying the target features rather than the whole training sample/patch. The effectiveness of the proposed data augmentation method is demonstrated through its application in extracting roads from the Swiss Siegfried map. This chapter addresses Research Question 2, and is presented in the paper by Jiao *et al.* (2022b).

Chapter 5 continues the research on the issue of generating training data. This chapter introduces a fast and effective deep learning approach for road extraction from historical maps by automatically generating training data through symbol reconstruction. Training data are generated by carefully crafting symbols on historical maps and applying those to contemporary geo-spatial vector data. Then, deep CNNs are applied for road extraction from the Siegfried map. Four training scenarios are designed for verifying the effectiveness and efficiency of the approach by comparing the results of manually labelled data and automatically generated data. This chapter tackles Research Question 3, and comprises the journal paper by Jiao *et al.* (2022a).

Chapter 6 further investigates road vectorization and classification from historical maps on the basis of road extraction results with deep learning in Chapter 5. The road extraction results from deep CNNs are pixel-based. Considering that the classified vector-based road data are more commonly used in many applications (e.g., urban planning and transportation), this chapter introduces a novel framework for vectorizing and classifying roads from historical maps based on deep learning and symbol painting. After obtaining pixel-wise raster road segmentations with deep learning, painting functions are defined for the road classes, which are able to paint each type of road symbol respectively. The proposed framework is capable of achieving road vectorization and classification simultaneously. It is validated through road vectorization and classification based on the Siegfried maps. This chapter answers Research Question 4, and is made up of the journal paper by Jiao *et al.* (2023).

Chapter 7 encapsulates the contributions and conclusions of this dissertation. In Section 7.1, a comprehensive summary of the results is presented, which is linked with the initial objectives highlighted in the introduction. A summary of the dissertation’s primary contributions is also presented. Following this, Section 7.2 delves into a discussion and offers recommendations for prospective research endeavours.

## 2. Background

### 2.1. A brief introduction to historical maps

The history of cartography aims at documenting development of maps since their beginnings, which can be dated back to the Stone Age (Crassard *et al.*, 2023). Ancient maps were engraved in stone, on clay tablets, bronze or wood or were painted on silk. Paper maps, which have been widely used for documenting geo-spatial information, emerged after the invention of paper (Wolodtschenko, A. and Forner, T., 2007; Utrilla *et al.*, 2009; Chiang, 2010). Paper maps were digitalized into raster format through scanning, which are the data source of almost all modern historical-map relevant studies (Heitzler and Hurni, 2020). Historical maps are cartographic representations of geographical features that were created in the past and specifically designed to convey information about (historical) events, geographic features, and cultural contexts of a particular time period (Chiang *et al.*, 2020a). Although historical maps can vary widely in their content and style, typically depicted features include human settlements, transportation infrastructures, hydrological networks, forests, and place names. (Leyk *et al.*, 2005). Historical maps play a crucial role in understanding the changes that have occurred over time, such as the evolution of political borders, the growth and decline of cities, or the transformation of the physical landscape (Casali and Heinemann, 2019; Saeedimoghaddam and Stepinski, 2020; Tschopp *et al.*, 2005). They can also provide insights into the historical context of events, and serve as valuable resources for historians, geographers, spatial planners and other researchers. Sometimes historical maps can be the only data source that provides relatively accurate spatial information in a certain period (Chiang *et al.*, 2020a).

In some cases, historical maps exist even in series, which are collections of maps that cover a specific geographical region or a country over a certain period, such as national map series. They are particularly useful in showing the changes and developments that occurred during that time.

#### 2.1.1. Historical Swiss maps

Historical Swiss maps include a set of successive map series that cover whole Switzerland.

- **Dufour Survey Map.** Dufour Maps are the first official topographic map series covering the whole of Switzerland. They were published between 1834 and 1863. The field survey maps were drawn at a scale of 1:25,000 for the Jura Mountains and the Swiss central plateau, and of 1:50,000 for the Alps. The final publication using copper engraving techniques was at a scale of 1:100,000 (Swisstopo, 2023). The maps provided for the first time a detailed representation of the country's landscape, settlements, transportation networks, hydrological features and political boundaries. They were initialized by Guillaume-Henri Dufour, who was a Swiss engineer, topographer, and army general, and were created by him together with his team of scientists, topographers and copperplate engravers. At that time, the traditional survey techniques of (graphical) triangulation and plane table mapping were utilized on the field (Imhof 1927; Loran *et al.*, 2018). The Dufour maps set a high standard for topographic mapping in Switzerland and laid the foundation not only for the nation-building of the new federal state of 1848 (Gugerli and Speich, 2002), but also for future map series, such as Siegfried maps and Old National maps.
- **Siegfried Map,** officially known as “Topographic Atlas of Switzerland”. In 1870s, many Dufour map sheets were already several decades old. Due to their small scale, Switzerland was only covered by a few synoptic map sheets (Grosjean, 2013). Thus, based on Dufour's large-scale field maps, Hermann Siegfried initialized the creation of Siegfried maps, which were more precise and offered a more detailed representation of Switzerland's terrain at a scale of 1:25,000 and 1:50,000. They were published between 1870 and 1949. Cartographers and surveyors used geodetic triangulation

together with graphical triangulation to measure distances, angles, and elevations. Then, the cartographic information was engraved and inked. The 1:25,000 scale maps were printed through copper engraving, and the 1:50,000 scale maps were printed using lithography (Swisstopo, 2023). In the 19th century, the mirror-inverted engraving on copper and limestone was among state-of-the-art technologies (Heitzler and Hurni, 2020). Apart from higher precision, another advancement in the Siegfried maps compared to Dufour maps was the use of contour lines to represent terrain and elevations. Moreover, Siegfried maps were updated every few years, which resulted in 3,903 sheets in total in the series.

- Old National map. The Old National map (at the time of its first publication called the “Neue Landeskarte = “New National Map”) followed the Siegfried maps from 1938 (1:50,000) and 1952 (1:25,000). They were updated in a regular 6-year cycle by the Swiss Federal Office of Topography (swisstopo<sup>6</sup>). Modern techniques such as photogrammetry and later optical remote sensing were utilized for the update of the sheets (Loran *et al.*, 2018). The National Map series comes in various scales to cater to different purposes, including 1:25,000, 1:50,000 and 1:100,000. Negative scribing on coated glass was used to create the printing plates, and then offset printing was applied for producing paper maps. Since the 2000s, the national map production is fundamentally modernised, leading to the development of the fully vector-based New National Map as the successor to the Old National Map (Hurni and Brandenberger, 2001, Swisstopo, 2023).

Figure 2.1 shows examples of Swiss historical maps listed above. They cover the same area located in north-west of Zurich city.

---

<sup>6</sup> <https://www.swisstopo.admin.ch/en/home.html>

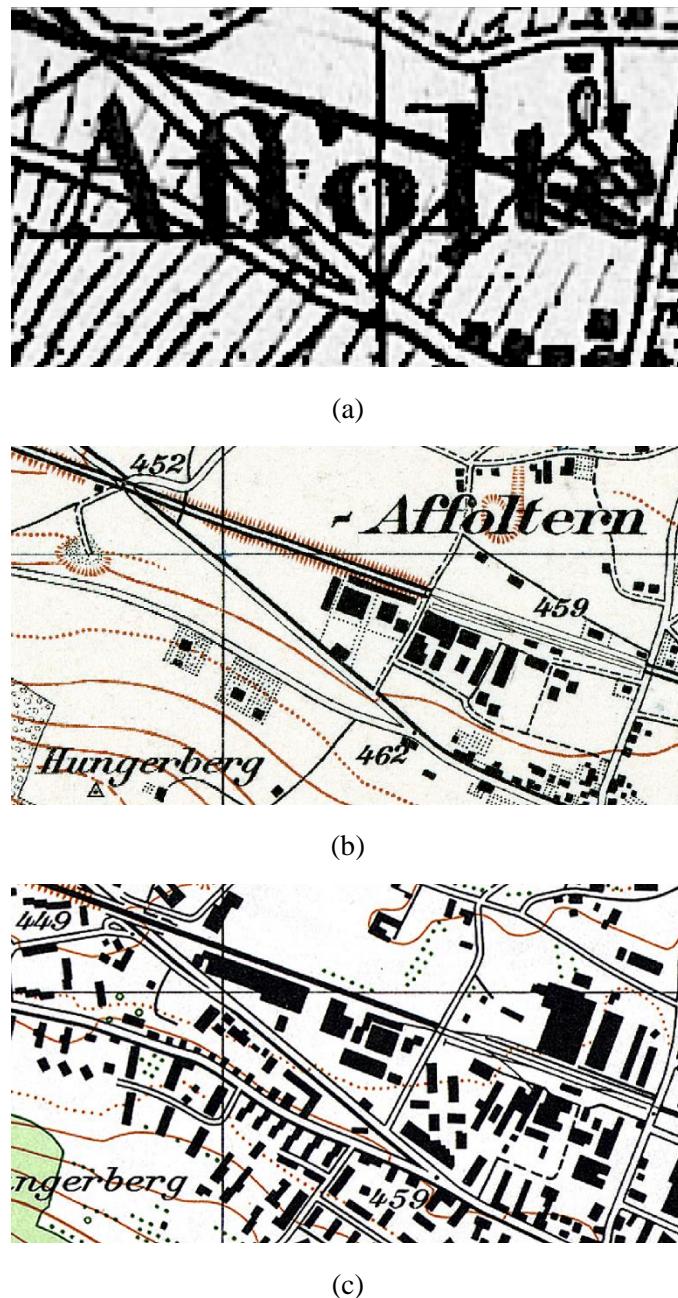


Figure 2.1. Examples of Swiss historical maps that cover a section located in north-west Zurich city. (a) The section on Dufour map (1:25,000), (b) the section on Siegfried map (1:25,000), (c) the section on Old National map (1:25,000). Geodata © Swisstopo

- There are other important historical map series in Switzerland. For example, the Gyger maps are a series of detailed hand-drawn topographic maps that cover parts of the Swiss cantons of Zürich, Aargau, and Thurgau. They were published in the mid-17<sup>th</sup> century at the scale of 1:32,000. Tools used for field measurements include “an astrolabe with a magnetic compass, a ruler and a paper with a triangular grid” (Wyder, 2006). The map series was named after its author Hans Conrad Gyger (1599-1674), a Swiss cartographer, painter, and engineer (Loran *et al.*, 2018). The maps were due to their high quality ahead of their time and were immediately classified for military reasons (Harley, 1988).

### 2.1.2. Selected foreign historical maps

- Ordnance Survey Maps. The British Ordnance Survey has been producing detailed maps of the United Kingdom since the 18<sup>th</sup> century (Rhind, 1991). The Ordnance Survey One-Inch Map lays a foundation to topographic maps produced in the English-speaking world. Originally, it was created for military use. It has gone through several editions, such as old series (1805-1874), third edition (1903-1913), popular edition (1919-1926), Second War revision (1940-1944) (Sebert, 2001; Haklay 2010).
- United States Geological Survey (USGS) Maps. They form an extensive archive of topographic maps that document the changing landscape of the US over more than a century (1884 to 2006). They cover the entire US with informative geographic features, such as contour lines, water bodies, buildings, vegetation, and transportation networks. This collection contains over 178,000 maps. They are scanned and georeferenced, thus are available as high-resolution digital images for researchers, historians, geographers, land-use planners, etc.
- Sanborn Fire Insurance Map. These series are a unique collection of large-scale maps that were created by the Sanborn Map Company<sup>7</sup>, primarily for assessing fire insurance risks in the US. The maps were produced in the US from the late 19<sup>th</sup> century to the mid-20<sup>th</sup> century. They typically had a scale of 1:600, which allowed for an impressive level of details, including the shape, size, use and materials, etc. of buildings, hydrants as well as the layout of streets and alleys in US cities and towns (Kodama, 2013). Different colours were used to distinguish different building materials. This detailed information was collected by hand on location (Lederle, 2017). The maps were updated every few years to account for changes in urban development and infrastructures. Additionally, the Sanborn Map Company also made maps for cities and towns in Canada and Mexico, thus resulting in a huge spatial coverage of more than 12,000 cities and towns (Mueller, 2004).

## 2.2. Digitalization, georeferencing and access of historical maps

In order to digitally preserve historical maps as well as to make them machine-readable, such paper maps are usually scanned and thereby converted into raster format, which facilitates studies for various purposes. In some cases, digitalization of historical maps is a requirement in the domain of cultural heritage. Preserving, analyzing and vectorizing historical maps has become trending in modern cartographic culture and thinking (Livieratos, 1998). Usually, line-wise scanning is employed to digitalize historical maps, through either contact or non-contact methods. Contact scanning is a lower-cost method suitable for originals that are not vulnerable to mechanical or chemical damage due to contact or illumination. Non-contact scanning, while more expensive, is generally less risky in terms of mechanical damage but still requires consideration of illumination impact during scanning (Tsioukas *et al.* 2006). There are several factors in the scanning procedure which affect the scanning results, including the scanning resolution which determines the level of detail captured in the digital image, illumination which can impact the final image quality and may potentially cause damage to the original map, colour accuracy which is crucial for creating a faithful digital representation of the map, file format and compression which affect the quality, size, and compatibility of the digital image, and the calibration of the scanner (e.g., colour calibration) (Daniil *et al.*, 2003; Tsorlini *et al.*, 2013).

Although high-resolution scanning can well preserve the graphical information of historical maps, the map images still cannot be directly used in Geographical Information Systems (GIS), be interpreted with modern image processing techniques and be compared with contemporary geo-spatial data, unless the map sheet is correctly and accurately georeferenced to its known or another reference system. Otherwise, if the map's reference system is unknown, the map should be adjusted to align with contemporary geo-

---

<sup>7</sup> <https://www.sanborn.com/>

spatial data through an optimal fitting procedure (Tsorlini and Hurni, 2015; Heitzler *et al.*, 2018; Luft and Schiewe, 2021). Moreover, incorporating location information into historical maps enables map holders (e.g., libraries, institutions) to index the archived maps by metadata, and makes the maps searchable online through spatial queries by the general public (Chiang, 2015). Additionally, comparing the spatial and semantic information of historical maps and other maps can be challenging, especially when they differ in age and style. Georeferencing makes them spatially consistent, thus enabling researchers to effectively compare information across maps (Höhn and Schommer, 2017; Schlegel, 2019; Luft and Schiewe, 2021).

Whether or not the spatial reference system of historical maps is known, it is necessary to find common points as control points on both the historical map and the reference map/data for georeferencing. The control points serve as equivalents, through which a link can be set up between these points and their locations in the real world. The georeferencing process can be modelled using various techniques, such as an affine transformation (Luft and Schiewe, 2021), a bilinear function (Hughes, 2012), a polynomial transformation (Chen and Tung, 2003) or more complex non-linear transformations such as rubber-sheeting (Rumsey and William, 2002), so that the original map image is warped to best fit the reference map or data. Then, the location of each pixel on the map image can be calculated or mapped to the real world. In order to achieve an accurate georeferencing, a sufficient number of suitable control points must be found. These control points should be easily identifiable, for example, intersections of grid lines, landmarks such as triangulation points, road intersections and churches. Manually selecting control points demands huge human effort, time and expert knowledge, considering that there are hundreds or thousands of map sheets covering a country or region (Chen *et al.*, 2008; Li and Briggs, 2012; Chiang *et al.*, 2014). One of the automatic methods developed to detect control points is template matching (Briechle and Hanebeck, 2001) together with random sample consensus (RANSAC) (Fischler and Bolles, 1981) to solve conflict matches (Luft and Schiewe, 2021). Another example is to apply Hough transform (Ballard, 1981) to detect grid lines and then localize their intersections as control points through a DBSCAN algorithm (Ester *et al.*, 1996) and bi-directional histograms (Heitzler *et al.*, 2018). Although georeferencing can be done largely in an automatic way, manual correction is still often required. The correction is of importance, as a single error may result in considerable displacements (Rumsey and William, 2002).

The reference data used as “standards” in georeferencing can be contemporary official survey data, geo-data (Chen *et al.*, 2008; Breunig and Zlatanova, 2011) and overhead imagery (Liu *et al.*, 2018). Interestingly, crowd-sourcing geo-spatial data such as OpenStreetMap (OSM<sup>8</sup>) data can also be used as reference data, which are publicly available around the world, but with still inhomogeneous coverage (Luft and Schiewe, 2021). In addition, web applications that allow for crowdsourced georeferencing by users are now available, such as Georeferencer<sup>9</sup>. Libraries and institutions that share historical maps with Georeferencer include the Library of ETH Zurich, the University Library Bern, the Central Library Zürich and the National Library of Scotland, etc. (Fleet *et al.*, 2012). One of the earliest examples of web-based georeferencing applications is MapRectifier, which allows users to upload and register maps and then georeference them (Kowal and Přidal, 2012). MapRectifier inspired the development of MapWarper<sup>10</sup>, which is designed for the use with OSM. In general, the essential workflow of an online georeferencing application typically consists of adding control points to a historical map, which have contemporary corresponding points. As long as a sufficient number of control points are provided, the historical map can be georeferenced or aligned to the contemporary reference data (Griffin and Lipkin, 2018).

Some historical maps are stored in libraries. For example, several collections of Swiss historical maps are available in the ETH library such as maps of the Canton of Zurich (ETH Library, 2023). The

<sup>8</sup> <https://www.openstreetmap.ch/>

<sup>9</sup> <https://www.oldmapsonline.org/partners/georeferencer/>, (accessed on 26.09.2023).

<sup>10</sup> <https://mapwarper.net/>

University Library Bern also has an extensive collection of historical maps (University Library Bern, 2023). Many Swiss historical maps and map series are kept and preserved by swisstopo. They are protected in “specially enclosed and air-conditioned premises” (swisstopo, 2023b). External users may hardly have access to these hard copy historical maps, as they can be very fragile. Exposure to air may damage the old paper. Thus, various online portals have been developed so that users can view and download historical maps. For example, GeoVITE (Geodata Versatile Information Transfer environment<sup>11</sup>) offers to Swiss universities and institutions the access to Swiss geodata, including Swiss national historical maps. It is operated by the Institute of Cartography and Geoinformation of ETH Zurich<sup>12</sup> and ETH Library (Tsorlini *et al.*, 2014; Gkonos *et al.*, 2019). Swisstopo offers a map viewer named “Journey through the time-maps” on its web portal<sup>13</sup>, which allows users to view and compare Swiss historical maps of different time periods. The USGS historical maps are available through the USGS National Map Viewer<sup>14</sup> and the USGS Historical Topographic Map Explorer<sup>15</sup>, and can be downloaded using the USGS US Topo and Historical Topographic Map Collection<sup>16</sup> (Chiang, 2010). The David Rumsey Map Collection<sup>17</sup> contains more than 150,000 maps of 16<sup>th</sup> to 21<sup>st</sup> century of North and South America, as well as maps of the World, Asia, Africa, Europe, and Oceania. The collection encompasses a variety of cartographic materials including school geographies, pocket maps, books of exploration, maritime charts apart from maps and atlases. The whole collection is housed at the David Rumsey Map Center in the Stanford University Library. There are now over 125,000 items online, and the website is updated regularly. Viewers can access not only high-resolution map images, but also compare and analyze the items. A vast collection of Sanborn maps can be searched, viewed and downloaded through an online portal of The Library of Congress<sup>18</sup> (Mueller, 2004). Willard Marriott Library of University of Utah (Arlitsch, 2002). An extensive collection of historical maps such as military survey maps and thematic maps can be accessed via Arcanum<sup>19</sup> (Biszak *et al.*, 2017). OldMapsOnline<sup>20</sup> combines collections of historical maps from multiple libraries and institutions worldwide, thus enabling users to search for and access historical maps (Southall and Pridal, 2012). A broader utilization of historical maps requires an efficient way to query and access historical maps, so that the maps can support researchers in wide domains (Chiang, 2015).

### 2.3. Quality issues in historical maps

Historical maps are usually subjected to quality issues, including positional errors or inaccuracy as well as vagueness and ambiguity in thematic information. Those errors can be identified by comparing objects on historical maps with their contemporary correspondence or with triangulation points (Crowell *et al.*, 1991). For example, georeferenced historical maps can be imported into MapAnalyst<sup>21</sup>, an online application for computing, visualizing and studying the planimetric accuracy of historical maps. The planimetric accuracy is computed as the root-mean-square error (RMSE) based on control points between the historical map and reference map (Jenny *et al.*, 2007).

---

<sup>11</sup> <https://geovite.ethz.ch/>, (accessed on 10.10.2023)

<sup>12</sup> <https://lkg.ethz.ch/en/>

<sup>13</sup> <https://map.geo.admin.ch/>

<sup>14</sup> <https://www.usgs.gov/programs/national-geospatial-program/topographic-maps>

<sup>15</sup> <https://livingatlas.arcgis.com/topoexplorer/index.html>

<sup>16</sup> <https://www.usgs.gov/us-board-on-geographic-names/domestic-names?p=262%3A1%3A0>

<sup>17</sup> <https://www.davidrumsey.com/>

<sup>18</sup> <https://www.loc.gov/rr/geogmap/sanborn/>

<sup>19</sup> <https://www.arcanum.com/en/>

<sup>20</sup> <https://www.oldmapsonline.org/>

<sup>21</sup> <http://mapanalyst.cartography.ch/>

### 2.3.1. Quality issues

#### 2.3.1.1. Production-relevant issues

Although historical maps were made by professional surveyors and cartographers, erroneous measurements still may exist. The surveying equipment used, which may at least be several decades old, can lead to positional inaccuracies. Geodetic or graphical triangulation methods may induce error accumulation (Savage and Burford, 1973). Mapping techniques such as generalized and abstracted data representation can also induce positional inaccuracy (Veregin and Hargitai, 1995; Burrough, 1996; Plewe, 2002; Levin *et al.*, 2010; Loran *et al.*, 2018). Moreover, as in all maps, the spatial scale of a historical map also indicates the (in)accuracy it can represent. A larger scale means higher capability of clearly presenting smaller geographical objects in detail, while under a smaller scale these small objects may be “neglected” due to generalization.

Some other positional inaccuracy is caused by the inherently fuzzy borders and transition areas of geographical features (Burrough, 1996). Examples are shown in Figure 2.2, in which (a) represents the wetland symbol, while (b) the symbol of open forest, (c) two wetland areas and (d) two open forests. Wetland is represented by clustered strokes, which are not connected with each other, although they form a contiguous texture (Jiao *et al.*, 2020). Open forest is symbolized as unevenly distributed circles and dots. The shape and size of the circles are somehow inconsistent. Wetlands and open forests shown in the figures have no distinct borders, making it hard or even impossible to determine their accurate spatial extents. These fuzzy borders and transition areas result in inaccuracy or vagueness in map interpretation. Actually, such vagueness may be induced in the surveying or mapping stage due to unclear semantic definitions used for identifying geographical features and their borders (Leyk *et al.*, 2005; Shi, 2009; Loran *et al.*, 2018).

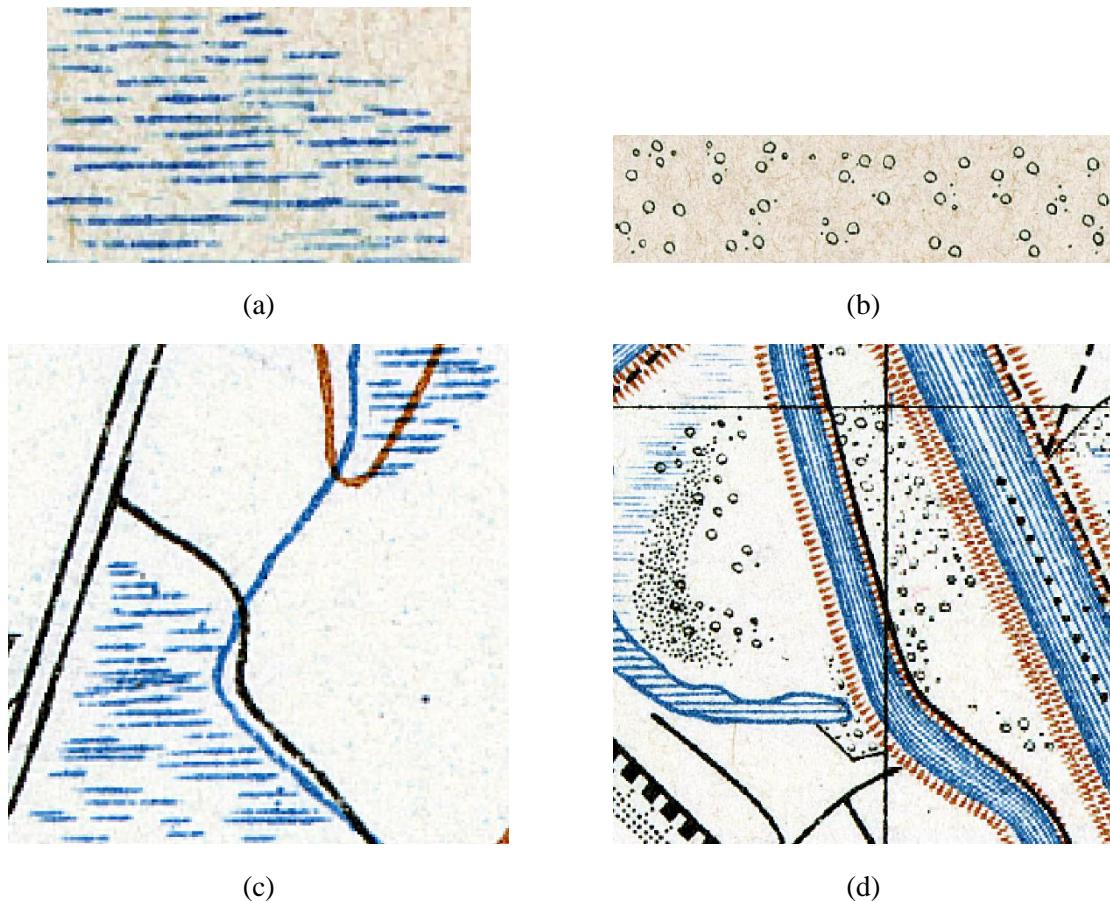


Figure 2.2. Examples of geographical features with fuzzy borders on the Siegfried map, (a) symbol of wetland, (b) symbol of open forest, (c) wetland areas and (d) open forests. Geodata © Swisstopo

The manual reproduction methods, such as copper engraving and lithography, can result in inconsistencies or heterogeneities of symbol representations. Figure 2.3 shows some examples, in which (a) displays a legend of road symbols: a dashed line representing a walking path (“Fussweg”), a single solid line representing a dirt road or mule track (“Feld- oder Saumweg”), a dashed line in combination with a solid line represents a driveway without reinforcement (“Fahrweg ohne Kunstanlage”), two parallel lines represent a reinforced road 3-5 meters wide (“Kunststrasse 3-5 Meter Breite”), a thin line together with a thicker line represent a reinforced road wider than 5 meters (“Kunststrasse über 5 Meter Breite”) and a thin line with a parallel thicker line with short strokes in between represent a tramway (“Strassenbahn”). Figure 2.3 (b) presents a “walking path”, a “dirt road or mule track” and a “driveway without reinforcement”, (c) shows a “reinforced roads 3-5 meters wide” and a tramway, and (d) depicts a “reinforced roads 3-5 meters wide”. It is observed in (b) that the lengths and widths of dashes as well as the gap between the dashes may vary for the same road class. Additionally, the width of the single solid line also varies. In (c), part of the tramway edges are represented with a thin line together with a thicker line, while the remaining part is shown with two parallel lines of the same width. Multiple “reinforced roads 3-5 meters wide” are shown in (d), but distances between the two parallel lines vary a lot.

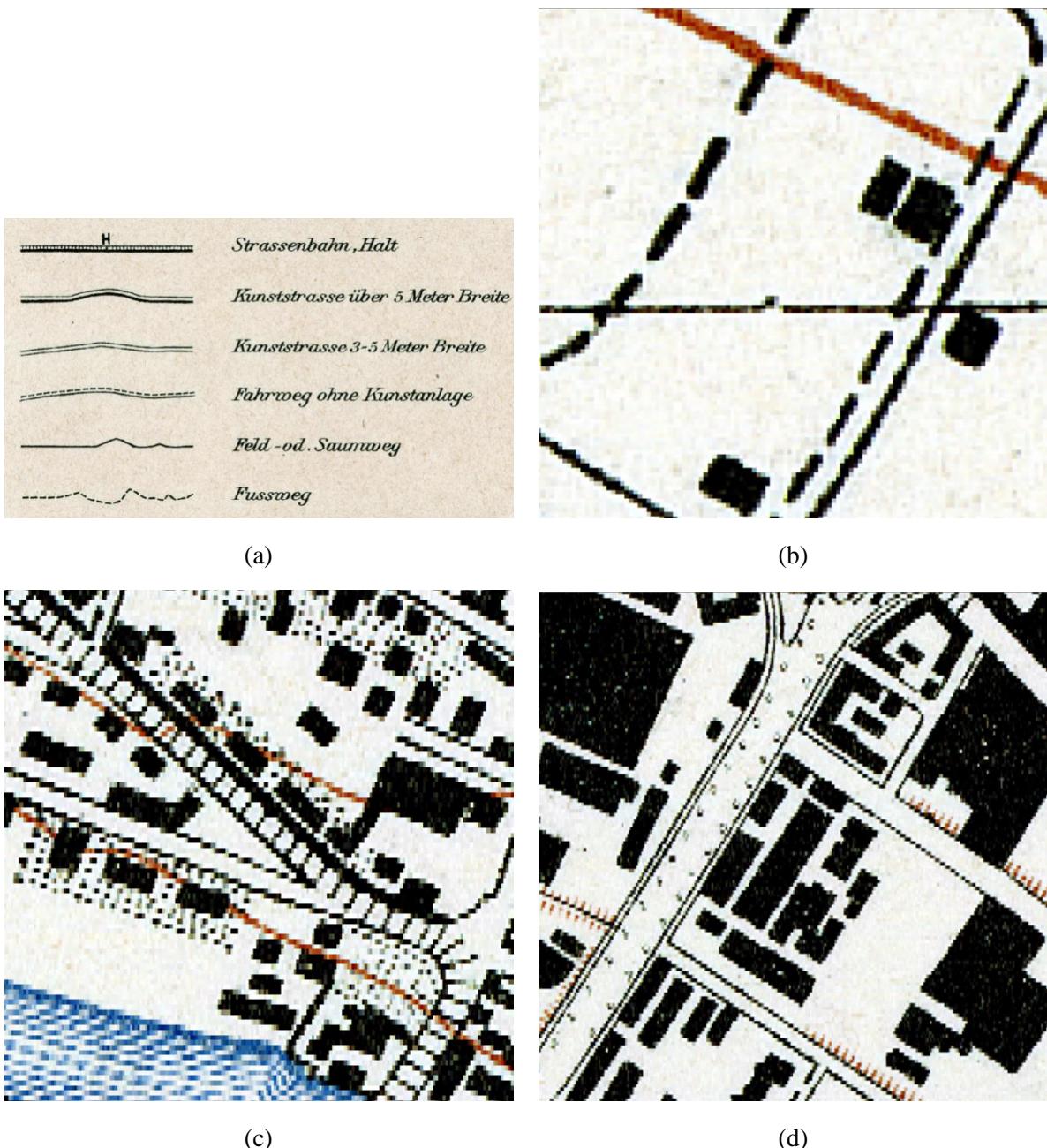


Figure 2.3. Examples of inconsistent symbol representations, (a) a legend of road symbols, (b) a “walking path”, a “dirt road or mule track” and a “driveway without reinforcement”, (c) a “reinforced road 3-5 meters wide” and a tramway, and (d) “reinforced roads 3-5 meters wide”. Geodata © Swisstopo

The inaccuracies mentioned above are a type of source error or the so-called production-relevant/-oriented uncertainties, which refer to the uncertainties inherent in historical maps (Muller, 1991).

### 2.3.1.2. Preservation-relevant issues

Positional distortion may also be induced during the long preservation period of historical maps. This distortion is due to chemical and physical deterioration of the paper maps (e.g., fold, bleaching, rips) caused by the preservation conditions (e.g., humidity, temperature) (Leyk *et al.*, 2005). The distortion results in shape and colour changes such as paper shrinkage (Liu *et al.*, 2019) or yellowing. Apart from these, there exist some “unknown pixels” due to dirt in the map image, which are marked with red arrows

in Figure 2.4. The yellow pixel cluster and the pink curve may be a result of inappropriate storage conditions.

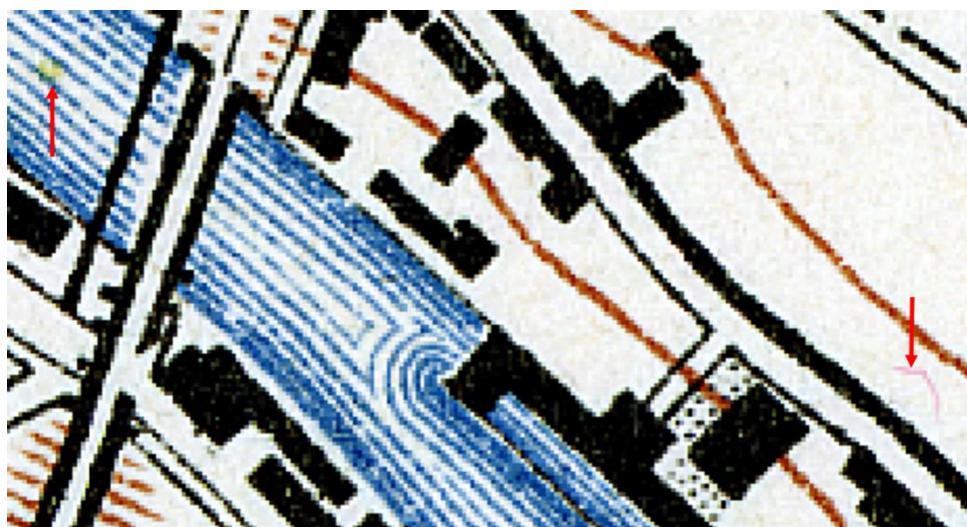


Figure 2.4. Examples of “unknown objects” pointed by arrows. Geodata © Swisstopo

### 2.3.1.3. Transformation-relevant issues

The media transformation process transforms paper-based historical maps into a digital format. Parameters set in the scanning process (e.g., scan resolution) also affect the quality of the digital map image. Scanning with higher resolution keeps more and accurate information, but requires more storage space, and potentially demands more image processing time. Higher resolution also means that unwanted dirt and spots (such as the example shown in figure 2.4) are scanned and kept in the digital image. Blurring, colour aliasing, chromatic aberration, false colour and mixed colour may be induced by scanning (Jiao *et al.*, 2021). The scanning process transforms the continuous colour distribution of a paper map to discrete pixel values in the digital image, thus colour sampling is inevitable, which results in blurring and colour aliasing. An example is shown in Figure 2.5. Optical misalignment of the scanner may induce chromatic aberration and false colour. The latter outputs an obviously different colour. The overlap of more than one colour in geographic elements from the paper map may cause the issue of mixed colours (Liu *et al.*, 2019), as demonstrated in Figure 2.5 (b). Figure 2.5 (a) presents a section from the Siegfried map, (b) and (c) display magnified image extracts of (a). From (b) and (c) we see that although road edges and forest edges should be black, some pixels around the edges are in dark green, dark blue or dark brown. Similar issues of blurring and colour aliasing exist in isolines which are represented with brown lines, as well as streams which are symbolized as blue lines. Moreover, the intersection between the blue stream and the brown isoline in Figure 2.5 (c) shows a mixed colour issue. The pixels on boundaries separating two distinct colours exhibit a difference from, yet simultaneously a resemblance to, the colours present on either side (Liu *et al.*, 2019). Figure 2.6 shows an example from the Swiss Old National map, where (a) presents a section of the map, and (b) shows magnified image extracts in the red rectangle in (a). It is observed that pixels around the boundary that separates the forest represented with light green colour and the hydrological feature shown as light blue colour exhibit mixed colours of light green, dark green and light blue. The situation is contradictory. Pixels around the forest border present a colour similar to, yet different from that of the forest on the left, while on the right similar to, but different from that of a water body. Moreover, pixels are in saw-toothed shape around boundaries of roads, forests, streams and isolines, etc. These issues arise from sampling in the scanning process, and may also be due to an inappropriate sampling approach during georeferencing. Additionally,

an issue of colour inhomogeneity exists in the background of the map, which may stem from bleaching issue in historical paper maps (Khotanzad and Zink, 1996; Khotanzad and Zink, 2003). Moreover, the applied georeferencing method and the selected parameter values also impact the positional accuracy of the map image. Table 2.1 summarizes the quality issues in historical maps and their possible corresponding reasons.

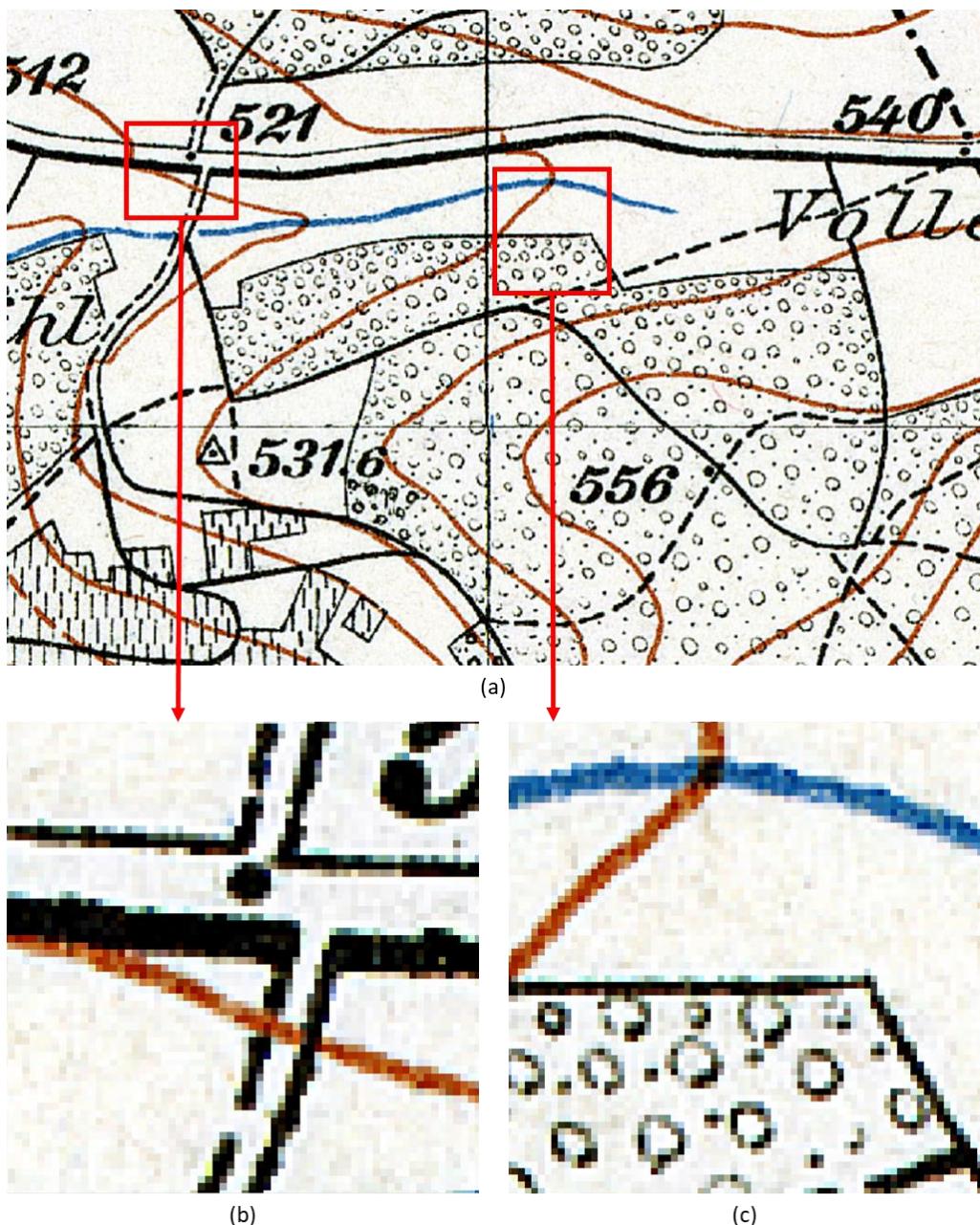


Figure 2.5. Examples of blurring, colour aliasing and colour inhomogeneity issues on the Siegfried map. (a) presents a section from the Siegfried map, (b) and (c) display magnified image extracts of (a). It is seen that road edges and forest edges show blurring and colour aliasing of dark green, dark blue or dark brown. Moreover, the overlap of the blue stream and the brown isoline results in mixed colour issue in (b). Geodata © Swisstopo

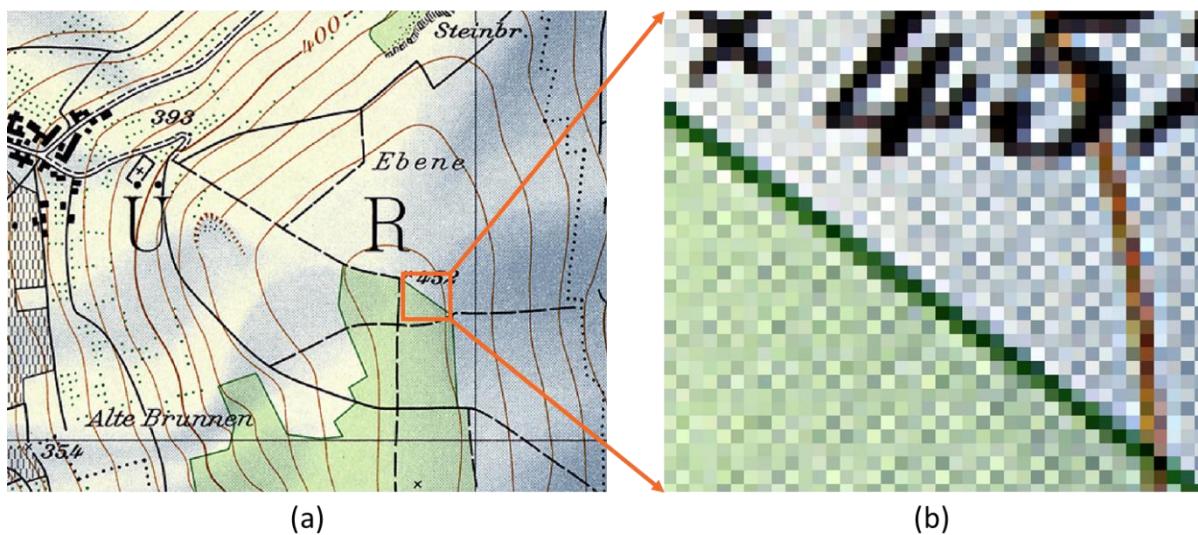


Figure 2.6. An example of colour aliasing around boundaries, (a) a subsection from Swiss Old National map, (b) magnified image extracts from (a). Geodata © Swisstopo

Table 2.1. Quality issues in historical maps.

Quality issues in historical maps	Possible reasons
Production-relevant issues	Surveying equipment, geodetic or graphical triangulation methods, mapping techniques, Spatial scale, Inherently fuzzy borders and transition areas of geographical features, Unclear semantic definitions used for identifying geographical features and their borders, Manual reproduction methods such as copper engraving and lithography
Preservation-relevant issues	Improper preservation conditions
Transformation-relevant issues	The scanning process, Georeferencing methods applied

### 2.3.2. Assessing and visualizing inaccuracies

The assessment of inaccuracies or distortions can be achieved by comparing the historical map with corresponding reference data/map of higher accuracy (Loran *et al.*, 2018; Schlegel, 2019). The assessments or estimation results are evaluated with a set of quantitative or even spatial measures, such as accuracy, percent correctly classified (PCC), and confidence interval (Livieratos, 2006). Accuracy measures the misrepresentation of map elements and can be quantitatively computed as RMSE. PCC evaluates to what degree a certain geographical feature class (e.g., forest) is mapped and represented correctly in the map by referencing the ground truth data. The ground truth data can be derived from other data sources such as data from test sites and present-day assessment. PCC is computed as the percentage of correctly classified pixels against all pixels. Confidence interval is used for evaluating the positional uncertainty along boundaries (Leyk *et al.*, 2005).

Distortions or errors can be graphically visualized by displacement vectors, a distortion grid and isolines of scale and rotation, etc. The displacement vector represents the difference between the location of a point identified on the historical map and the reference map. A vector line originates at a point on the historical map and terminates at the location of its corresponding point on the reference map. Extremely

long vectors indicate large positional errors in the historical map. It is easy to visually identify possible patterns by checking local displacement vectors (e.g., direction, length) in a specific area (Symington *et al.*, 2002; Jenny and Hurni, 2011). Figure 2.7 is taken from Tsorlini *et al.* (2013), presenting the distortion in Zurich on the georeferenced Gyger Map against the contemporary reference map. Displacement vectors are shown with blue arrows. A distortion grid visualizes the local deformation and rotation of a historical map, represented by rotated, compressed or enlarged meshes. For a long time, map historians have been using distortion or deformation grids; they were originally made manually (Wagner, 1896; Forstner, 1998). With computer-assisted methods, distortion grids can be generated based on displacement vectors, but different algorithms may generate different grid shapes depending on the applied techniques, such as different interpolation techniques, various regression methods and digital simulation of manual construction techniques, etc. (Weis, 1985; Tobler, 1994; Beineke, 2001). When dealing with distortion in a local area of a historical map with relatively high accuracy, a distortion grid may have limited capacity to visually present such deformation, as local distortions are difficult to visually identify (Jenny and Hurni, 2011). Isolines of scale connect locations with the same scale, while isolines of rotation link locations with equal rotation. Other means of computing and visualizing map distortions include triangular nets, differential distortion analysis, and topological inconsistency. For triangular nets, spatial differences between the historical map and the reference map can be visualised by changing the thickness of edges or colours of faces of triangles. Differential distortion analysis takes the coordinates of the ground control points on the reference map and the historical map as input, and computes distortions for points on the historical map. Topological inconsistency refers to the topological difference between the historical map and the reference map (Vervust *et al.*, 2018; Reckziegel *et al.*, 2021). The quantitative analysis of the distortion or inaccuracy can potentially revert the production process of a historical map. The analysis is also useful to determine the applicability of a historical map for contemporary questions. For example, Loran *et al.* (2018) demonstrate that historical maps from the mid-19<sup>th</sup> century onward usually have an acceptable level of accuracy for analyses of forest cover.

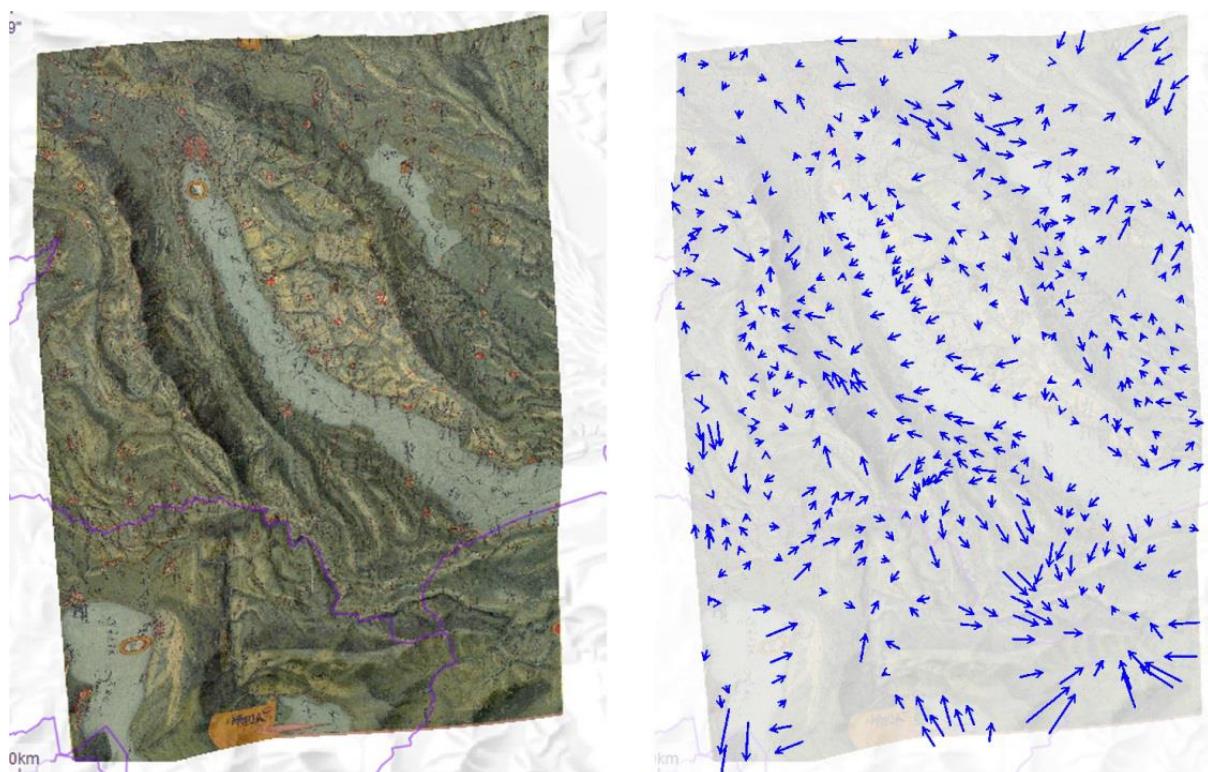


Figure 2.7. A Figure from Tsorlini *et al.* (2013), showing the distortion in Zurich on the Gyger Map (1:32,000) of 1660 compared to the same area in modern maps. Displacement vectors are represented with blue arrows.

## 2.4. Segmentation and vectorization of historical maps

In geosciences, historical maps serve as an indispensable data source for extracting various geographical features, such as roads, buildings, water bodies, forests, etc. Although the focus of this thesis is put on road networks, in this section, we summarize the segmentation and vectorization of common geographical features from historical maps. Segmentation and vectorization of historical maps are usually prerequisites while applying them to various fields. They divide a historical map into multiple more meaningful and easier-to-analyze parts or segments based on certain characteristics (e.g., colour, texture).

### 2.4.1. Road and railway extraction

Road networks together with railroads represent a tangible documentation of the development of our civilization (Jacobson, 1940; Uhl *et al.*, 2022). They embody the sequential process of human settlement, urban sprawl and city infrastructure accessibility (Axhausen *et al.*, 2011). Thus, they have attracted notable attention in the domain of segmentation and vectorization of historical maps.

A number of studies have been conducted to extract road networks from various historical maps. For example, Perret *et al.* (2015) reconstruct road networks at the French national level from the historical map of Cassini in the 18<sup>th</sup> century based on a collaborative methodology. Concretely, the digitalization of the maps is implemented using a shared PostgreSQL<sup>22</sup> database. The operators are able to remotely digitalize data on the same database simultaneously using GIS editing tools (e.g., QGIS<sup>23</sup>) in combination with a WMTS (Web Map Tile Service) layer supplied by the National Institute of Geographic and Forest Information (IGN<sup>24</sup>) as background. Saeedimoghaddam and Stepinski (2020) extract road intersections from USGS historical maps with a region-based deep convolutional neural network (RCNN). Road intersections are particularly useful for the registration of geo-spatial data from different sources, e.g., overhead imagery and raster maps (Chen *et al.*, 2008; Liu *et al.*, 2019). Ekim *et al.* (2021) extract five road classes from the Deutsche Heereskarte 1:200,000 Türkei (DHK 200 Turkey) using U-Net++ (Zhou *et al.*, 2018), aiming at reconstructing the past transportation networks. A similar study is reported by Avcı *et al.* (2022), in which roads are extracted and classified with deep learning methods from DHK 200 Turkey. The authors examine and compare the performance of road extraction with U-Net++, Deeplabv3 (Chen *et al.*, 2017) and an attention network (Niu *et al.*, 2021). The extracted roads can be used for investigating transportation conditions in the past and for spatio-temporal analysis of historical events and human activities. In another study, contemporary geographical data of roads, railroads and water lines are aligned to their correspondence on USGS historical maps, and these aligned data can be used as training data for vectorizing these features (Duan *et al.*, 2017; Duan *et al.*, 2020). Similarly, roads are extracted from USGS historical maps with contemporary road segments as auxiliary data (Uhl *et al.*, 2022). The extracted roads can be used for a comprehensive understanding of the evolution of transportation infrastructures. Jiao *et al.* (2022) use contemporary Swiss geo-spatial data (e.g., roads, vegetation, building footprints) to automatically create training data for the vectorization of roads from the Swiss Siegfried map. This paper will be elaborated in Chapter 5 of this dissertation. Chiang *et al.* (2020a and 2020b) test a set of CNNs for extracting railroads from USGS historical maps.

---

<sup>22</sup> <https://www.postgresql.org/>

<sup>23</sup> <https://qgis.org/en/site/>

<sup>24</sup> <https://www.ign.fr/institut/identity-card>

#### 2.4.2. Label extraction

Label extraction and recognition are helpful for constructing spatial query of historical maps, for the study of toponymy and for creating a dataset of named geographical features (e.g., named roads) (Chiang and Knoblock, 2015). Various approaches have been proposed and developed for label detection and recognition. For example, Chiang and Knoblock (2015) develop a semi-automatic method to detect and recognize labels from historical maps of Baghdad. First, the text layer is separated based on colours of the text examples provided by the user. Then, texts are rotated to the horizontal direction and a commercial optical character recognition (OCR) product named ABBYY FineReader 10<sup>25</sup> is employed for text recognition. Schlegel (2021) develops an end-to-end workflow to automatically detect and recognize labels from historical maps of Hamburg, Germany. Bounding boxes of labels are first detected from the map using a deep learning-based text detection method named Strabo<sup>26</sup> (Chiang and Knoblock, 2013). The bounding boxes are unified according to certain criteria (e.g., the distance between centroids of bounding boxes). Then, the open-source engine Tesseract OCR<sup>27</sup> is used for text recognition. The recognized labels will be potentially useful for an automatic comparison between historical maps of different periods or different map series. In the study by Weinman *et al.* (2019), a neural network constituted of ResNet (He *et al.*, 2016) and convolutional layers is trained for detecting bounding boxes of map labels from U.S. historical maps selected from David Rumsey Map Collection (Ray *et al.*, 2018). Then, a bidirectional long short-term memory (LSTM) network is used for text recognition (Sherstinsky, 2020). Moreover, a synthetic data generation method is developed, which automatically produces training samples.

#### 2.4.3. Contour line extraction

Extraction of contour lines from historical maps contributes to the creation of digital elevation models (DEM) and reconstruction of the 3D landscape of the past. Chen *et al.* (2006) apply the colour key set technique (Khotanzad and Zink, 2003) and gray-level histogram information to extract contour line pixels from scanned historical maps. As a novelty, the authors propose and develop a local window segmentation method based on line-tracing to refine the contour lines. Ghircoias and Brad (2011) develop a semi-automatic procedure to vectorize contour lines from scanned historical maps. The layer of contour lines is separated using colour image segmentation (CIS) techniques, and the colour of contour lines is indicated by a user. The layer is then binarized and the contour lines are thinned and pruned with morphological operations. Then, the contour lines are vectorized and a user has to label the elevation for each line. A 3D DEM is reconstructed based on the contour lines. Likewise, Brüngger (2011) extracts and vectorizes contour lines from the Swiss Old National Map (1:50,000), which is then used for constructing DEM of glaciers. Wiesmann (2016) vectorizes contour lines from the Swiss National Maps 1:50,000 and generates the historical elevation models of glaciers.

#### 2.4.4. Human settlement extraction

Long-term historical building data are essential for the analysis of the development of human settlements, such as morphological homogenization of settlements, the interaction between the settlements and natural environment (Räth *et al.*, 2023). Uhl *et al.* (2017) extract buildings from USGS historical maps by employing contemporary building data as a spatial context layer to guide the sampling process of

<sup>25</sup> <https://www.abbyy.com/company/news/abbyy-finereader-10-breaks-new-ground-in-document-digitization/>, (accessed on 23. 09. 2023)

<sup>26</sup> <https://github.com/spatial-computing/strabo-text-recognition-deep-learning>, (accessed on 26.09.2023)

<sup>27</sup> <https://github.com/tesseract-ocr/tesseract>, (accessed on 26.09.2023)

building graphics from historical maps. In this way, thousands of building samples are automatically generated for training. Then, the samples are fed into the CNN-architecture “LeNet” (LeCun *et al.*, 1998) for training. Heitzler and Hurni (2020) introduce a machine learning-based workflow for building footprint vectorization from Swiss Siegfried map. Buildings are first segmented using an ensemble of U-Net. Then they are vectorized through contour tracing and orientation-based clustering, which achieve high-quality building footprint representations of perpendicular corners and parallel walls. The obtained building footprint polygons are made publicly accessible to the scientific community of Switzerland by integrating them into GeoVITE.

#### 2.4.5. Hydrological feature extraction

Hydrological features with explicit spatial information in the past allow researchers to study and reveal the spatio-temporal dynamics of the hydrological and ecological situation. Jiao *et al.* (2020) extract wetlands from the Swiss Siegfried map and vectorize the wetlands using the Geo-spatial Data Abstraction Library (GDAL<sup>28</sup>). Xia *et al.* (2022) employ CNN-based template matching to extract wetlands from Swiss Old National Maps. The template matching models are adapted from quality-aware template matching (QATM), which takes into account the quality of the matches by assigning a confidence or similarity score to each match (Cheng *et al.*, 2019). Wu *et al.* (2022) segment hydrological features including streams, rivers, wetlands and lakes from Swiss Siegfried maps by integrating atrous spatial pyramid pooling (ASPP) (Chen *et al.*, 2017) to U-net. ASPP enables the network to capture image features and context at multiple scales.

### 2.5. Applications of historical maps

Historical maps are an invaluable documentation and source of retrospective geographical information of natural features and human activities over long time periods. They are sometimes the only source of long-term geo-spatial data, which benefit research and applications in various domains (e.g., ecology, toponymy) (Chiang, 2017; Casali and Heinemann, 2019). The retrospective information facilitates a detailed interpretation of the evolution of the landscape, human settlements, land cover, etc. (Herold, 2015; Uhl *et al.*, 2018). The comparison between historical maps as well as between historical maps and contemporary data is essential for the investigation of land-use and land cover changes (Schlegel, 2019). A simple visual comparison can be done with online tools. An example is “A journey through time<sup>29</sup>” offered by swisstopo (Rickenbacher, 2013). By adjusting map transparency, historical maps and contemporary maps that cover the same region can be compared intuitively. Another example is “Mapjunction<sup>30</sup>”, which was originally developed for the city of Boston’s planning agency. Two maps from different years are overlaid and a user compares them using the 4-way slider to adjust the boundary and transparency. A detailed and comprehensive comparison of maps from different periods reveals detailed and quantitative results of land-use changes. San-Antonio-Gómez *et al.* (2014) compare historical maps, including the 1775 Domingo de Aguirre map and the 1835 Plano general del pueblo, with orthophotos of 2005 to analyze urban and landscape changes of the Real Sitio de Aranjuez, Spain. Levin and Kark (2023) compare 130 historical maps and aerial photos to quantitatively reconstruct land cover and land use changes over 250 years on Norfolk Island, Australia. Examples of these historical maps include the Arrowsmith’s 1840 map<sup>31</sup> (Skurnik, 2020) and the Australian Geographic map of Norfolk Island (1990). Weber *et al.* (2020) reconstruct specifically glacier inventories of 1882-1916 in

<sup>28</sup> <https://gdal.org/>

<sup>29</sup> <https://www.swisstopo.admin.ch/en/maps-data-online/maps-geodata-online/journey-through-time.html>

<sup>30</sup> <https://www.mapjunction.com/>

<sup>31</sup> <https://www.asmp.esrc.unimelb.edu.au/>

Nordland of northern Norway from Norwegian historical gradteigskart maps<sup>32</sup>. More importantly, they compare the reconstructed glacier extents with existing glacier inventories of 1976, 1988 and 2000 in order to quantify glacier recession in Nordland.

Moreover, historical maps are employed to realistically build 3D streetscapes of the past for education and entertainment. Google creates “time travel” experience through reconstructing 3D cities from crowd-sourced historical maps and photos using deep learning. Sánchez-Berriel *et al.* (2021) implement an interactive 3D application based on the Web Graphics Library (WebGL<sup>33</sup>) using the oldest known map of the city San Cristóbal de La Laguna<sup>34</sup>, Spain, which was drawn in 1588 by a military engineer. A user can have an immersive experience of the city back to the 16<sup>th</sup> century. This map is also used by Pérez Nava *et al.* (2023) to create 3D representation of the city. Immersive experiences can be created with virtual reality (VR) technology. Additionally, EPFL<sup>35</sup> and the University Ca’Foscari of Venice launch a project aiming at developing a 4D model for Venice, named the Venice Mirror World<sup>36</sup> using geo-historical datasets derived from historical documents including maps. Results of the project were presented in prestigious museums throughout the world. The Venice Mirror World is part of the Time Machine Europe<sup>37</sup> project.

## 2.6. Techniques for historical map image processing

Historical map processing is a field at the intersection of computer vision, GIS, and digital cartography. Various digital image processing as well as geodata processing techniques have been used for enhancing, analyzing and interpreting historical maps (Chiang *et al.*, 2014). The motivation of historical map image processing is to derive and preserve unique historical information to better understand the complex landscape and its changes over long time. It’s aiming at generating either image (raster) or vector data of geographic and cartographic features from historical maps for subsequent use in various fields.

### 2.6.1. Rectification and resampling

Before feature recognition and spatial analysis of historical maps, the first step is to georeference them. The georeferencing process involves map image rectification and resampling. Rectification assigns the image a spatial reference system or transforms the image to a spatial reference system. Resampling interpolates the pixel values for the new pixel configuration (Affek, 2013). Examples of map image rectification techniques include affine transformation and polynomial transformation, etc. Nearest neighbor interpolation, bilinear interpolation and cubic interpolation are common techniques for image resampling. Although historical maps are assigned or transformed to a target reference system via rectification and resampling, the process may induce inaccuracy. Also, different georeferencing methods may lead to different results, i.e., a point from the historical map may have different coordinates in a spatial reference due to different georeferencing methods.

### 2.6.2. Colour image segmentation

One common way to extract features from raster map images is to group the pixels with similar colours. Colour image segmentation (CIS) is used to divide a colour image into meaningful and visually distinct

<sup>32</sup> <https://kartverket.no/om-kartverket/historie/historiske-kart>

<sup>33</sup> <https://get.webgl.org/>

<sup>34</sup> <https://turismo.aytolalaguna.es/en/torrianis-street-map/>

<sup>35</sup> <https://www.epfl.ch/en/>

<sup>36</sup> <https://www.epfl.ch/research/domains/venice-time-machine/>

<sup>37</sup> <https://www.timemachine.eu/>

regions or segments based on colour information, thereby obtaining geographic features (Cheng *et al.*, 2001). The basic intention of CIS is to identify colour-homogeneous regions, which is essential and critical in image processing and pattern recognition tasks. Here are some typical CIS techniques used in historical map image processing.

#### 2.6.2.1. Histogram thresholding

Histogram thresholding is a simple yet effective technique in CIS. It operates under the premise that images consist of distinct features with varying colour ranges. Spatial contiguity is not explicitly taken into account in histogram thresholding (Chiang *et al.*, 2014). In this method, the histogram of an image is used to assess and manipulate the distribution of colours in the image based on a graphical representation of colour frequency (Novak and Shafer, 1992), which can be divided into multiple peaks, each representing a specific colour cluster or feature. A threshold value, typically corresponding to a valley between two neighboring peaks, can be identified to separate the adjacent classes (Cheng *et al.*, 2001). As colour images, historical maps contain multiple features or classes. Thus, multiple histogram thresholding can be employed to divide colour space by thresholding each component histogram to obtain different classes. Histogram thresholding has been widely used for historical map image segmentation (e.g., Leyk and Boesch, 2010; Mello *et al.*, 2012). In addition to segmentation, histogram thresholding is also used for filtering of historical map images (e.g., Gobbi *et al.*, 2019).

Histogram thresholding has its advantages and disadvantages in map image processing. The advantages lie in simplicity, computational efficiency, and visualization (Tan and Isa, 2011). Also, it allows for a visual and intuitive way to segment images by choosing a threshold value based on the histogram. The disadvantages include challenges in threshold selection, limited to bimodal histograms, and the need for post-processing, etc. Histogram thresholding works best when the image histogram is bimodal (e.g., two distinct peaks corresponding to object and background classes). For images with several objects whose intensity ranges overlap, problems can occur if a global threshold value is used. (Medina-Carnicer *et al.*, 2009). Additionally, post-processing steps are usually needed to refine the segmented results, especially in the presence of noise or artifacts.

#### 2.6.2.2. Colour space clustering

Colour space clustering is an image processing technique used to group similar colours in an image into clusters or segments, which can be regarded as the multidimensional extension of the concept of thresholding (Cheng *et al.*, 2001). It is an unsupervised learning approach to identify classes of objects without any prior knowledge. How to assign each pixel to a certain class or cluster depends on distance measures in the colour space of an image. Existing algorithms that have been commonly used in colour space clustering include K-means that clusters image (colours) into K distinct groups based on colour similarity, C-means, fuzzy C-means, Mean-Shift that shifts pixel values towards the “mean” within a local region used for colour clustering and image smoothing, etc. (Fukunaga and Hostetler, 1975; Lloyd, 1982; Chiang *et al.*, 2014). These colour space clustering techniques have been applied to historical maps for image segmentation, such as line features extraction (Henderson and Linton, 2009; Yang *et al.*, 2012), forests extraction (Herrault *et al.*, 2013), urban area extraction (Uhl *et al.*, 2021), etc. Figure 2.8 presents an example of applying Mean-Shift to the Siegfried map, where (a) shows the original map, and (b) the output. It is observed that noise on buildings and the background is largely removed. But details such as gardens represented with black dots distributed in array are “blurred”.

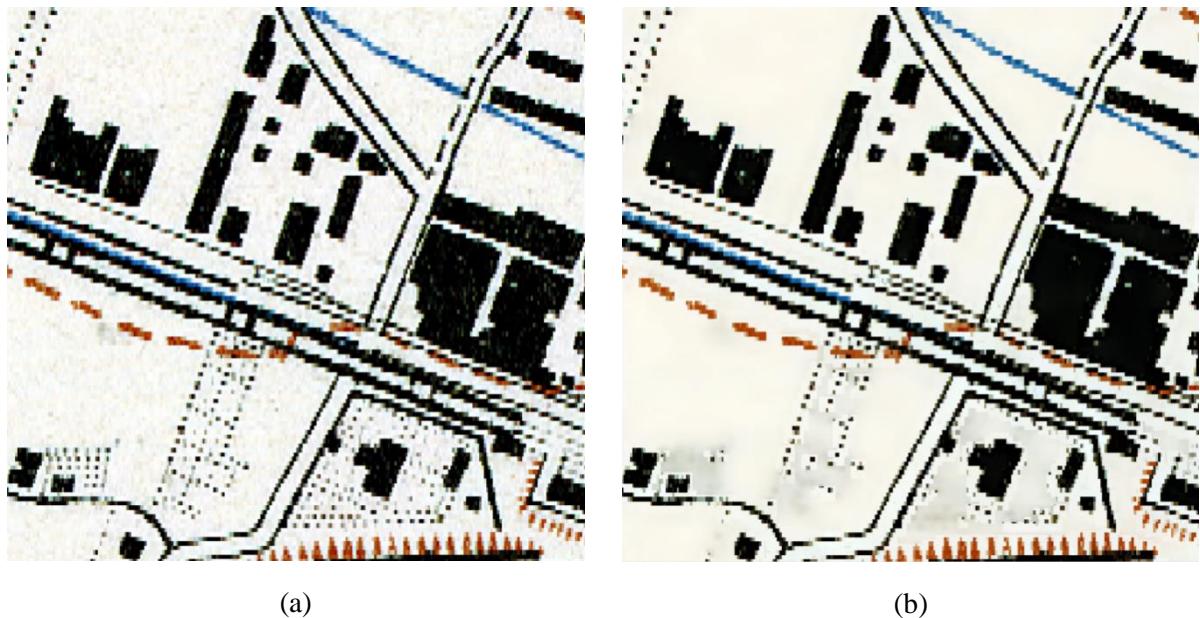


Figure 2.8. An example of applying Mean-Shift to a section of a Siegfried map sheet, (a) the original map, (b) the result processed with Mean-Shift. Geodata © Swisstopo

Its capability of segmenting objects or features based on colour information makes it useful for historical map image processing tasks. Nonetheless, how to determine the appropriate number of clusters ( $K$ ) is a crucial step and can be challenging. Different values of  $K$  can lead to different segmentation results. In addition, post-processing is required sometimes to refine the segmentation results.

### 2.6.3. Morphological operations

Mathematical morphology has been demonstrated to be useful for processing binary and grayscale images (Serra, 1983). In image processing, morphological operations refer to a set of mathematical operations to analyze and manipulate the shapes and structures within images. These operations are particularly useful for tasks such as image segmentation, noise reduction, edge detection, skeletonization, and object identification (Comer and Delp, 1999). Morphological operations are conceptually defined and implemented by applying a structuring element (a moving window or kernel) to the input image and modifying pixel values based on their spatial relationships with neighboring pixels. The two fundamental morphological operations are dilation and erosion, which increase/decrease the size of objects in an image by expanding/shrinking their boundaries respectively. Additionally, they can be combined to perform more complex operations like opening and closing. Opening is a sequence of erosion operations followed by dilation, used for removing noise from an image while preserving the overall shape of objects. Closing is a sequence of dilation operations followed by erosion, used for filling small gaps in objects and connecting nearby objects. Morphological operations have been applied to historical maps for various applications, such as road extraction (Dhar and Chanda, 2006; Chiang *et al.*, 2014; Jiao *et al.*, 2022a), buildings detection (Heitzler and Hurni, 2020; Chazalon *et al.*, 2021) and contour line extraction (Liu *et al.*, 2019).

Morphological operations have shown advantages in image processing, such as high efficiency and simplicity. However, their effectiveness depends on careful parameter selection and moving window design. Moreover, morphological operations can be computationally expensive in certain cases when dealing with large images (Landini *et al.*, 2019).

#### 2.6.4. Template matching

Template matching is an image processing technique used for finding or locating a specific template image within a larger image (Brunelli, 2009). In most image analysis problems, it is often necessary to measure similarity between two or more different objects or images. Template matching employs measures of similarities between objects based on certain mathematical algorithms (Hashemi *et al.*, 2016). Concretely, the template of the target feature is slid over the map image and a distance value is simultaneously computed between the template and the corresponding area in the image, which is used to identify the areas that best match the template. Distance metrics used for template matching include cross-correlation, squared differences and correlation coefficient, etc. (Lewis, 1995). Template matching can be particularly useful for tasks in historical maps, such as object recognition (Budig, 2018; Chazalon *et al.*, 2021), and georeferencing (Höhn *et al.*, 2013; Luft and Schiewe, 2021). Figure 2.9 shows an example of four templates used for wetland extraction from the Siegfried map.

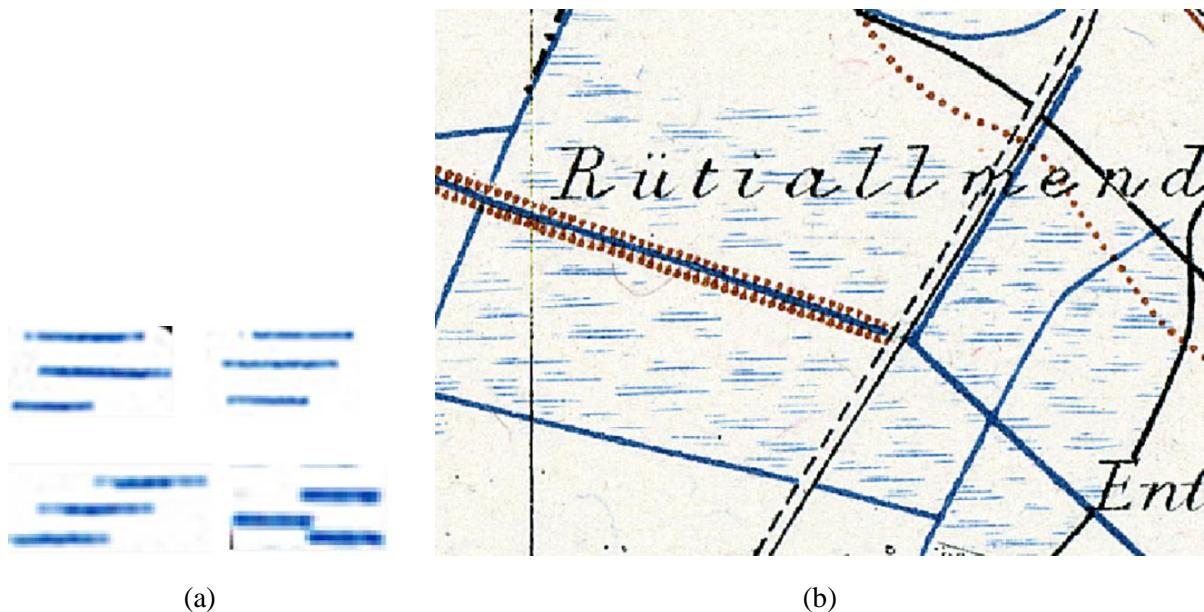


Figure 2.9. An example of template matching for wetland extraction, (a) four wetland templates, (b) a section from the Siegfried map. Geodata © Swisstopo

Template matching is conceptually simple in image processing and analysis since complex models or training data are not required. It can easily detect different types of objects just by changing the template without tedious training procedures. However, there are also some limitations to this technique, such as sensitivity to variations and lack of robustness. It implies that template matching lacks scale-invariance, rotational-invariance, and intensity-invariance (Han, 2021). Consequently, alterations in the template's shape, dimensions, orientation, or colour, as well as changes in the characteristics of the target objects, can impact its ability to successfully identify the desired objects.

#### 2.6.5. Optical character recognition

Optical character recognition (OCR) is a technology used in image processing to recognize and extract text from images or scanned documents (Hamad and Mehmet, 2016). Map labeling has specific underlying semantics and follows cartographic rules. The labels can have varying orientations or overlap

with other geographical elements (Chiang *et al.*, 2014). Hence, text extraction and recognition from historical maps is not as straightforward as in the case for contemporary images.

OCR has been used to extract text from historical maps (Chiang, 2017; Schlegel, 2021), which is a step-by-step procedure (Thorat *et al.*, 2022). First, pre-processing is required to make the text more recognizable, such as removing noise or converting image into binary format, without missing any significant information, which can largely improve the effectiveness of OCR. Second, text segmentation is applied, which is aiming at separating text from the image background. Next, classification and recognition are conducted to divide the separated text into different classes. The used techniques include statistical techniques, neural networks, and kernel methods. Last, post-processing is sometimes required to refine the OCR results. Nowadays, a number of OCR platforms or software are accessible, such as Google Docs OCR, Tesseract OCR, ABBYY FineReader, Transym, I2OCR, etc. (Thorat *et al.*, 2022).

Applying OCR to historical maps can unlock valuable toponymy information and make these cartographic heritages searchable and accessible for research and education purposes. However, it is also necessary to acknowledge the limitations of OCR for historical map image processing. On the one hand, the accuracy of text recognition can vary depending on the map condition, typography and language. On the other hand, human expertise remains crucial for ensuring the quality of OCR results (Hamad and Mehmet, 2016).

#### 2.6.6. Machine learning

Machine learning, within the realm of computer science and artificial intelligence, is a field dedicated to the development of algorithms and models that enable computer systems to acquire knowledge, recognize patterns, and make data-driven predictions or decisions without explicit programming (Mitchell, 1997). It draws inspiration from various scientific disciplines, including statistics, mathematics, and cognitive science, to create computational systems that can automatically learn and improve from experience. At its core, machine learning involves the utilization of large datasets to train algorithms, allowing them to uncover underlying patterns, relationships, and dependencies within the data. The potential of machine learning in image processing is vast and continues to expand rapidly. It provides powerful tools and techniques, thus is playing a pivotal role in image classification, feature extraction and image segmentation (Camps-Valls *et al.*, 2011).

Deep learning, a subfield of machine learning, has emerged as a powerful and transformative approach for solving complex problems across diverse scientific domains in recent years. Deep learning focuses on the development and utilization of neural networks with multiple layers, known as deep neural networks. A method called backpropagation is used to compute gradients, which makes it feasible and efficient to train multilayer architectures, so weights of neurons can be optimized. Deep neural networks are designed to automatically learn and represent complex patterns and features from data, making them well-suited for various scientific applications. Unlike traditional machine learning methods, deep learning excels at automatically extracting hierarchical features and patterns from raw data (LeCun *et al.*, 2015).

Among neural networks, a convolutional neural network (CNN) is a regularized network that learns features by itself via filter (or kernel) optimization (Chellapilla *et al.*, 2006). CNN is designed to process data composed of multiple arrays (e.g., a colour image constituted of three 2D matrixes). A deep CNN consists of multiple layers, i.e., convolutional, pooling, and fully connected layers. Convolutional layers apply convolutional operations with a sliding filter (or kernel) to the input data, thereby extracting features, such as edges, textures, and patterns. The role of a convolutional layer is to detect local features. Pooling layers are used to reduce the spatial dimensions of the feature maps produced by the convolutional layers. A max pooling layer computes the maximum of a local patch in a feature map, while an average pooling layer calculates the average. Pooling helps reduce the computational

complexity of the network and makes it more robust to variations in input size and position. Similar to traditional neural network layers, a fully connected layer connects every neuron in one layer to every neuron in another layer. It is applied after convolutional and pooling layers to make final predictions/classifications based on the extracted features (Sahu and Dash, 2021). Activation layers (or functions) are applied to the output of convolutional and fully connected layers. They enable the gradients to flow backward through the network in back propagation. Activation functions introduce non-linearity into the network, thereby boosting its capacity to model complex and non-linear relationships in data. Frequently used activation functions include Rectified Linear Unit (ReLU) (Nair and Hinton, 2010) and Exponential Linear Unit (ELU) (Clevert *et al.*, 2015). ReLU is partially linear but not fully linear. It clips negative values in the input and transfers positive values to the output, while ELU models the negative values in the input with exponential function (LeCun *et al.*, 2015; Bawa and Kumar, 2019). Figure 2.10 presents a simple yet typical architecture of a deep CNN. Training a deep CNN usually requires large amounts of data. Due to its superior performance in image processing and analysis, machine learning, especially (deep) CNN has been widely applied to historical maps, such as road extraction (Chiang *et al.*, 2020a; Jiao *et al.*, 2022b), building reconstruction (Uhl *et al.*, 2017; Heitzler and Hurni, 2020; Farella *et al.*, 2021), water body extraction (Jiao *et al.*, 2020; Wu *et al.*, 2022; Xia *et al.*, 2022), label extraction (Laumer *et al.*, 2020), and geographic entity alignment across different historical maps to build knowledge graphs (Sun *et al.*, 2021).

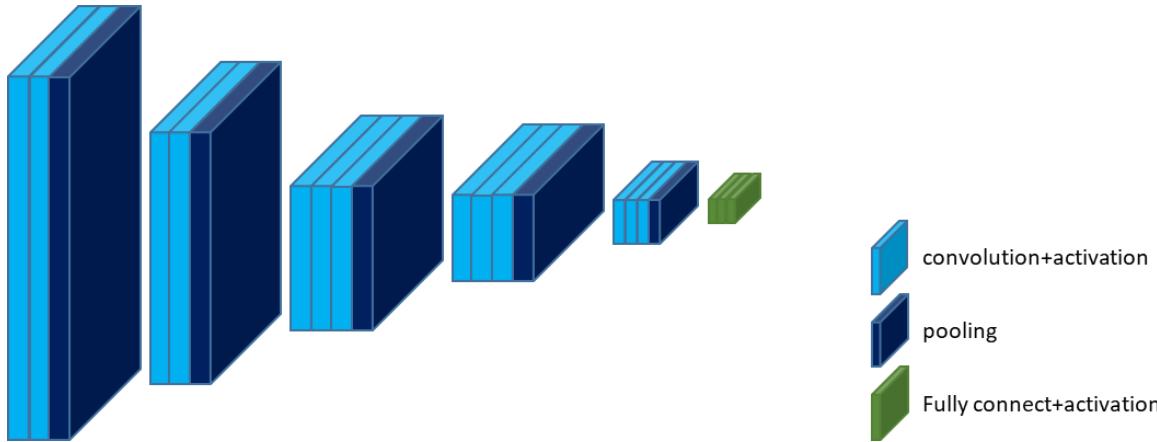


Figure 2.10. A simple yet typical architecture of a deep CNN consisting of convolutional layers, max pooling layers and fully connected layers.

Apart from CNN, Reinforcement Learning (RL) is also used in historical map-related studies. RL is designed to address problems that can be modeled with an agent interacting with an environment in order to achieve a goal. In RL, the agent is the learner who takes actions based on its current state and the information it has about the environment. The environment is the domain or context in which the agent operates. The state is a representation or observation of the current situation in the environment. The action is the set of possible moves or decisions that the agent can take in a given state. An action leads to transitions from one state to another. After each action, a reward is provided to the agent, which is used to indicate the immediate benefit or punishment of the agent's action. The goal of the agent taking actions is to maximize the cumulative reward in a sequence of actions (Sutton and Barto, 2018). RL is used by Duan *et al.* (2020) to model the process of aligning contemporary vector data (e.g., roads, railways) to its correspondence in historical USGS maps, thereby automatically creating training data for subsequent feature extraction tasks.

Application of Generative Adversarial Networks (GAN) (Goodfellow *et al.*, 2020) is also seen in historical map processing. It has become a powerful tool for generating synthetic data, such as images

and text. The key idea behind GAN is to train two neural networks, named the generator and the discriminator, in a competitive framework. Initially, the generator produces random synthetic data that is of low quality and can be easily distinguished from real data. The discriminator provides feedback to the generator by trying to correctly classify the real and synthetic data. It provides gradients that guide the generator to improve its output. Then, the generator uses this feedback to adjust its parameters and produce more convincing synthetic data or more similar data to the real data. This back-and-forth training process continues iteratively, with the generator gradually being able to generate increasingly higher-quality synthetic data that can hardly be distinguished by the discriminator. GAN is used by Li *et al.* (2019) to automatically generate a large amount of training data for text detection from historical maps, and Christophe *et al.* (2022) for historical map style transfer.

Although CNNs achieve promising performance in image classification, semantic segmentation and object detection, a kernel in convolutional layers mainly focuses on the core region of interest, thus may ignore context information. Some methods are proposed and implemented to tackle this issue such as atrous spatial pyramid pooling (ASPP) (He *et al.*, 2019) and Multi-receptive-field CNN (Liu *et al.*, 2020), but still they are not capable of considering all relevant input features. The attention mechanism, which is originally developed for natural language processing, allocates different weights to different elements in the input, so that output is generated by focusing on the most relevant elements of the input (Niu *et al.*, 2021). The self-attention mechanism, which operates on the input to discover and model relationships between all elements in the input, learns to pay “attention” to specific elements. Transformer is the name of an architecture that relies entirely on a self-attention mechanism to draw global dependencies between input and output without using convolutional layers. Usually, a Transformer consists of steps including input embedding, positional encoding, self-attention, layer normalization, feed-forward and output (Vaswani *et al.*, 2017). Li *et al.* (2022) propose a novel spatial language model based on the principle of Transformers to detect and link text from historical USGS maps.

#### 2.6.7. Raster-to-vector conversion

Raster data sometimes need to be converted to vector formats including points, (poly)lines and polygons, which are more suitable for spatial analysis, performing geoprocessing operations (e.g., buffering, intersection), editing and updating, data integration, data sharing, and data compression (Congalton, 1997; Liu *et al.*, 2017). Before and in the 1990s, manual or semi-automatic raster-to-vector conversion was popular, in which an operator traced the boundaries of geographical features (Eikvil *et al.*, 1995). Later, automatic line tracing methods were developed to trace either the medial lines or boundaries from the map image (Suzuki and Yamada, 1990; Liu, 2002). Heitzler and Hurni (2020) employ contour tracing to vectorize segmented buildings from the Siegfried map. Edge detection techniques, such as Canny operator (Canny, 1986) and Sobel operator (Sobel, 1968), can be used to extract boundaries from historical maps. Before the use of edge detectors, an image thresholding step can be necessary to separate foreground and background pixels (Brown *et al.*, 2012). The boundaries then can be vectorized using the OpenCV Library (Bradski, 2000). The performance and accuracy of raster-to-vector conversion depend on the techniques applied, parameter values selected, etc.

A specific task for processing historical maps (e.g., image classification, feature extraction) usually requires the use of multiple techniques. As the focus of this thesis is road extraction from historical maps, a comprehensive and detailed review of road extraction techniques from raster maps is expounded in Chapter 3.



### 3. A survey of road feature extraction methods from raster maps

---

The following chapter presents the journal paper “A survey of road feature extraction methods from raster maps” by Chenjing Jiao, Magnus Heitzler, and Lorenz Hurni, <https://onlinelibrary.wiley.com/doi/full/10.1111/tgis.12812>, published on 15 August 2021 in *Transactions in GIS*. Content-wise only figural examples of raster maps in the Appendix are added as complement to the published version of the paper in the journal. Apart from this only formatting changes and corrections of spelling have been made.

---

#### Abstract

Maps contain abundant geospatial information, such as roads, settlements, and river networks, to name a few. The need to access this information to carry out analyses (e.g., in transportation, landscape planning, or ecology), as well as advances in software and hardware technologies, have driven the development of workflows to efficiently extract features from maps. The aim of this article is to provide a comprehensive overview of such methods to extract road features from raster maps. The methods are categorized based on the classes of techniques they employ (e.g., line extraction), as well as their subclasses (e.g., line tracing, Hough transform), the amount of user intervention required (e.g., interactive, automatic), the required data (e.g., scanned maps, contemporary vector data) and the produced results (e.g., raster-based predictions, vector-based results, attributes). Additionally, recent road extraction methods from overhead imagery, together with evaluation methods that will possibly benefit road extraction from raster maps, are reviewed. Furthermore, the evolution of this research field is analyzed over the past 35 years and the limitations of the current techniques, as well as possible future directions, are discussed.

#### 3.1. Introduction

Nowadays, a large number of scanned raster maps of different ages are available in digital map archives. Yet the wealth of cartographic information, such as roads, building footprints, hydrography, and so forth, is still locked in these maps and therefore cannot be directly analyzed or used in GIS (Chiang *et al.*, 2020; Chiang *et al.*, 2011). Therefore, there is a high demand for effective and efficient feature extraction methods.

Cartographic information on road data has a particularly broad application domain. Long-term historical road data are used to analyze the evolution of the road networks. For instance, the development of the Swiss road and railway network during the period of 1950–2000 is investigated by Erath *et al.* (2009). Swiss road-based accessibility since 1,850 is reported by Axhausen *et al.* (2011) to understand the impact of transport investment on the economy and land use patterns. Strano *et al.*, (2012) reported the analysis of the evolution of the road networks for almost 200 years in a large area located north of Milan. Masucci *et al.* (2014) studied the growth of London's street networks over 224 years by modeling the networks in dual representation and analyzing their topological properties, such as degree, connectivity, average path length, and so forth. Zhao *et al.* (2015) statistically analyzed the evolution of the OpenStreetMap (OSM) road network for Beijing. Casali and Heinemann (2019) studied the growth of the road network from 1955 to 2012 in Zürich, Switzerland. Wang *et al.* (2019) analyzed the evolution of the road network in Changchun, China from 1912 to 2017. Road data extraction from historical maps is a prerequisite of the studies on road evolution that cover a long time span. Furthermore, historical

road data are also used to realistically reconstruct streetscapes of the past for education, entertainment, and research purposes. Moreover, especially in geoscience, one application is geospatial data integration as road features exist across various geospatial data sources. For example, maps and remote sensing imagery covering the same area can be aligned through utilizing the extracted common road intersection points from these two spatial data sources as control points or “glue” (Chen *et al.*, 2008; Chen *et al.*, 2004). Chiang *et al.* (2009) use road intersection templates containing the information of the positions, connectivity, and orientations of the road intersections extracted from raster maps to extract roads from remote sensing imagery, so that these two data sources are integrated. In addition, the extracted road features can be applied for improving the extraction results themselves. For instance, in the study by Chiang *et al.* (2009), a template is constructed based on each extracted road intersection. Subsequently, the localized template matching (LTM) method (Chen *et al.*, 2008) is utilized to adjust the intersection points to the precise location, thereby improving the accuracy of the extracted road intersections in terms of position, connectivity, and orientation. Another study reported by Chiang and Knoblock (2009a) used the identified road pixels to generate a road template to find road pixels that cannot be detected by the Hough transform (Ballard, 1981). Furthermore, up-to-date road data are essential to update the existing road database, including fitting the existing road data to the real landscape, improving the planimetric accuracy, and deriving the height of the road centerlines (Eidenbenz *et al.*, 2000; Fortier *et al.*, 2001; Zhang, 2003), and timely road maps are crucial in applications including disaster management, urban planning, car navigation (Baltsavias and Zhang, 2005; Itonaga *et al.*, 2003), intelligent transportation systems (Zhang *et al.*, 2005), and impervious surface extraction (Wang *et al.*, 2015).

Despite the wide application domains of road data, the extraction of road features from raster maps can be challenging due to the similarity of road symbols to those of other features (e.g., isolines, streams), the long length of some features (e.g., highways spanning whole countries), the adjacency of other map elements with the same colour (e.g., buildings), and the marginal differences between the symbols representing different classes of roads (e.g., main roads, country roads) (Herold, 2015). Especially paper maps that have been printed decades or even centuries ago may suffer from poor quality. The reasons might be inaccurate printing technologies as well as chemical and physical deterioration (e.g., bleaching, fractures, paper distortion). Furthermore, blurring and colour aliasing may be induced by the scanning process (Leyk *et al.*, 2005; Liu *et al.*, 2019). Additionally, map readers interpret the maps by checking the map context, such as road labels or map legends, which is a challenging task for machines (Chiang *et al.*, 2009). To overcome these problems, researchers have developed numerous approaches, such as line tracing, morphological operation, colour image segmentation (CIS), machine learning, and so forth.

This article aims to review the studies on road feature extraction and vectorization from raster maps, as well as analyzing the development and progress of these road extraction techniques. Also road extraction methods, as well as evaluation metrics from overhead imagery, that will possibly benefit the task for raster maps are reviewed, aiming to bring inspiration from the field of remote sensing. Both interactive and automatic road extraction methods from raster maps are reviewed in this article. Although there already exist surveys on feature extraction and vectorization from topographic maps (Chiang *et al.*, 2014; Liu *et al.*, 2019), they aim to cover map feature extraction methods in general and thus lack information on the special case of road feature extraction, such as interactive road extraction methods. Interactive or semi-automatic methods refer to the process in which a user collaborates with a computer. Interactive methods take advantage of a computer's ability to precisely delimit a feature and to combine it with a user's high-level understanding of the map image (Moučka, 2018). Originally, interactive methods worked on paper maps or binary raster map images, and roads were extracted together with other linear features (e.g., counter lines) (Eikvil *et al.*, 1995; Kennie, 2014; Stevenson, 1994). Recently, interactive methods have generated promising road extraction results from colour map images and have the capability to segment road features with other linear features. Until now, interactive road extraction has still played an important role in data acquisition. However, in comparison with interactive methods, automatic approaches nowadays clearly dominate the field of road extraction from raster maps. Automatic methods largely reduce manual intervention during the extraction process. But they require the road features to be consistent within the data source. Moreover, a set of rules or parameters usually must be pre-defined before the automatic extraction and vectorization of road features (Stevenson, 1994).

Although the degree of automation and accuracy of feature extraction has been continuously improving in recent decades, fully automatic extraction of road features still cannot be achieved because human inspection of the raster maps is necessary to achieve reliable map interpretation results (Bin and Cheong, 1998; Chiang *et al.*, 2014; Suzuki and Yamada, 1990; Yang *et al.*, 2012). For example, an interactive correction step is necessary to extract unrecognized and remove falsely recognized features. Apart from this, in the existing surveys, the evolution of road extraction techniques is rarely investigated. Thus, a brief analysis of the development and progress of road extraction techniques is given to review the research trends over the past decades, as well as to indicate possible future research directions. Furthermore, a novel categorization scheme is proposed to classify the methods, based on the concrete techniques applied and their purposes. In this categorization scheme, detailed technique characteristics of each method are presented. Classical and general techniques for line extraction (e.g., Hough transform, morphological operations), image segmentation (e.g., histogram technique, K-means), noise filtering (e.g., conventional filters like mean-shift), and so forth are listed in Table 3.1 and expounded in the corresponding text. The remainder of this article is structured as follows. Section 3.2 first shows the categorization scheme, and reviews road extraction methods from raster maps. Selected road extraction methods and evaluation methods from overhead imagery are reviewed in Section 3.3. Section 3.4 reports the analysis of the development and progress of the road extraction methods, followed by a discussion of the current technical limitations and future technology trends. Section 3.5 concludes.

## 3.2. Road feature extraction from raster maps

### 3.2.1. Categorization

This section presents a detailed review of the road extraction methods based on raster maps. The methods are categorized in a novel hierarchical categorization scheme shown in Table 3.1. The methods are first characterized and classified according to the applied techniques, whether user intervention is required, the input data, and the results they produce. Furthermore, the techniques are classified based on their goals, including line extraction, image filtering, and CIS. To clearly show the procedure of each method, numbers are used to indicate the orders of the techniques applied in the method. Whether user intervention is required is indicated by “i” (interactive) or “a” (automatic) (e.g., “i, 1” means the technique is the first one in the method and is used in an interactive way; “a, 2” means the technique is used automatically and secondly). The methods in the table are ordered according to publication year, so the evolution and development trend of the methods can be explicitly shown, based on which the development and progress of the road extraction methods is analyzed in Section 3.4. Methods with the same techniques and technique orders, as well as the same input and output, are clustered and shown in the same colour.

Table 3.1. Categorization of road extraction methods based on raster maps.

Paper	Line extraction			Image filtering			CIS			Input	Output
	Line tracing	Parallel characteristic extraction	Hough transform	Conventional filters	Morphological operations	CNN	Histogram technique	K-means	Other clustering techniques	SVM	
Miyatake (1985)	a <sup>†</sup> , 2*	a, 1								Raster maps	Road centerlines
Nakajima <i>et al.</i> (1985)	a, 2	a, 1								Raster maps	Road centerlines and the connectivity of intersections
Amin and Kasturi (1987)	a, 1			a, 2						Raster maps	Road centerlines
Suzuki <i>et al.</i> (1987)	a, 1									Raster maps	Road centerlines
Alenany and Kasturi (1988)	a, 1				a, 2					Raster maps	Road centerlines
Kasturi and Alenany (1988)	a, 1				a, 2					Raster maps	Road centerlines
Nagao <i>et al.</i> (1988)		a, 1								Raster maps	Road centerlines
Kasturi <i>et al.</i> (1989)	a, 1				a, 2					Raster maps	Road centerlines
Nagao <i>et al.</i> (1990)		a, 1								Raster maps	Road centerlines
Suzuki and Yamada (1990)	a, 1									Raster maps	Road centerlines
Yamada <i>et al.</i> (1990)							a, 1			Raster maps	Road pixels
Yamada <i>et al.</i> (1991)							a, 1			Raster maps	Road pixels
Yamada <i>et al.</i> (1993)							a, 1			Raster maps	Road pixels
Ebi <i>et al.</i> (1994)							a, 2	a, 1		Raster maps	Road centerlines

Paper	Line extraction			Image filtering			CIS			Input	Output
	Line tracing	Parallel characteristic extraction	Hough transform	Conventional filters	Morphological operations	CNN	Histogram technique	K-means	Other clustering techniques	SVM	
Stevenson (1994)	i, 1									Raster maps	Centerlines
Eikvil <i>et al.</i> (1995)	i, 1									Raster maps	Centerlines
Ahn <i>et al.</i> (1997)				a, 2				i, 1		Raster maps	Road centerlines
Bin and Cheong (1998)		a, 1								Raster maps	Road centerlines
Nishijima and Watanabe (1998)		a, 1								Raster maps	Road centerlines
Watanabe and Oshitan(2001)		a, 1								Raster maps	Road centerlines
Yin and Huang (2001)				a, 1			a, 2			Raster maps	Road centerlines
Liu (2002)	a, 4	a, 3		a, 1	a, 2					Raster maps	Road centerlines
Itonaga <i>et al.</i> (2003)					a, 1					Raster maps	Road centerlines
Chiang <i>et al.</i> (2005)	a, 3	a, 2				a, 4		a, 1			Raster maps
Chiang and Knoblock (2006)										a, 1	Raster maps
Dhar and Chanda (2006)	a, 3	a, 4			a, 2			a, 1			Raster maps
Chiang and Knoblock (2008)	a, 3	a, 2				a, 4		a, 1			Raster maps
Chiang <i>et al.</i> (2009)	a, 3	a, 2				a, 4		a, 1			Raster maps

	Line extraction			Image filtering			CIS			Input		Output	
Paper	Line tracing	Parallel characteristic extraction	Hough transform	Conventional filters	Morphological operations	CNN	Histogram technique	K-means	Other clustering techniques	SVM			
Chiang and Knoblock (2009a)		a, 4	a, 1	a, 5			a, 3	a, 2		Raster maps, user labels	Road pixels		
Chiang and Knoblock (2009b)		a, 4	a, 1	a, 5			a, 3	a, 2		Raster maps, user labels	Road centerlines, intersection positions		
Henderson and Linton (2009)							a, 1				Road pixels, intersections positions		
Henderson <i>et al.</i> (2009)							a, 1	a, 2			Raster maps	Road pixels	
Leyk (2009)									a, 1		Raster maps	Road centerlines	
Linton (2009)				a, 2		a, 1				Raster maps	Road centerlines, intersections positions		
Pezeshk and Tuttwiler (2010)					a, 1						Raster maps	Road centerlines	
Pezeshk (2011)					a, 1						Raster maps	Road centerlines	
Pezeshk and Tuttwiler (2011)					a, 1						Raster maps	Road centerlines	
Callier and Saito (2011)			a, 1			a, 2					Raster maps	Road pixels	
Chiang <i>et al.</i> (2011)			a, 2	a, 3							Raster maps, user labels	Road centerlines	
Yang <i>et al.</i> (2012)	i, 3			a, 2			a, 1				Raster maps	Centerlines	
Chiang and Knoblock (2013)	a, 5	a, 3	a, 1	a, 4				a, 2			Raster maps, user labels	Road centerlines, intersection positions and orientations	
Henderson (2014)					a, 2		a, 1	a, 1			Raster maps	Road centerlines, intersections positions	

Paper	Line extraction			Image filtering			CIS			Output	
	Line tracing	Parallel characteristic	Hough transform	Conventional filters	Morphological operations	CNN	Histogram technique	K-means	Other clustering	SVM	
Kennie (2014)	i, 1										Paper maps
Duan <i>et al.</i> (2017)								a, 1			Raster maps, road vector data
Chiang <i>et al.</i> (2020)							a, 1				Raster maps
Duan <i>et al.</i> (2020)									a, 1		Raster maps, road vector data
Saeedimoghaddam and Stepinski (2020)					a, 1		a, 2				Raster maps

† “i” refers to an interactive way to implement the technique, while “a” an automatic way.

\* The numbers refer to the orders in which the techniques are applied.

Line extraction directly yields lines and usually operates in the image space. It includes the techniques of line tracing, parallel characteristic extraction, and Hough transform. Image filtering yields pixel-based results and carries out operations in the image space. The image space is defined by the height and width of the image as well as its channels. The channels typically contain three components, one for each of the primary colours red, green, and blue. Sometimes an alpha channel as a fourth component is present.

Image filtering-based road extraction techniques encompass conventional filters, morphological operations, and convolutional neural networks (CNNs). The term “conventional” is used to distinguish from CNNs. A CNN is considered an image filter-based technique since its convolutional layers, which are probably the most important layer type, essentially make use of a multitude of image filters to modify an input image.

CIS yields pixel-based results and operates in the colour space. The colour space consists of the following dimensions: one for each colour red, green, and blue. Sometimes an alpha channel is added as a fourth dimension. In rare cases, these dimensions are transformed into an alternative colour space, such as  $L^*u^*v$ . The width and height of the image do not play a role in the colour space, which distinguishes it from the image space. CIS-based road extraction techniques comprise the histogram technique, K-means, other clustering techniques, and support vector machine (SVM).

Moreover, the input and output of each method are also included in this categorization scheme. As input to the methods, paper maps refer to hard copies of printed maps and raster maps refer to maps that exist in raster form, which is typically obtained via scanning. As for the output, if a method only extracts road pixels without a subsequent vectorization step, the output in the table is “road pixels.” Otherwise, if a method is specifically developed for extracting road centerlines in vector form, the output is “road centerlines.” But if the method is developed to extract centerlines of all kinds of linear features (e.g., roads, contour lines, streams), its output is “centerlines.” In this way, a detailed and comprehensive characterization of the methods is presented and the techniques are explicitly associated with their goals.

### 3.2.2. Line extraction-based road extraction

#### 3.2.2.1. Line tracing

##### 3.2.2.1.1. Interactive line tracing

A large portion of the graphical elements on maps is composed of line features. Thus, researchers have implemented many approaches to extract and vectorize linear features, including roads (Kennie, 2014; Stevenson, 1994). Interactive line tracing vectorizes linear features through following the lines and recording line points every few pixels. It is a popular and intuitive approach that gives the operator an opportunity to guide the process and utilize human knowledge (Eikvil *et al.*, 1995).

In line tracing, the operator interactively identifies a section of the scanned raster map image, zooms into the selected area, places the screen cursor on the linear feature, and determines the direction of the line. Then, the software takes over and automatically traces the linear features by following and drawing along the lines.

The  $x$  and  $y$  coordinates of points on the line are recorded at a preset interval. When some complicated features are reached, such as the end point of a line or a potential intersection, the program waits for operator input. A potential intersection is recognized by a sudden increase in line width. Some line tracing programs have the capability for dashed line recognition, line intersection straightening, line generalization, and so forth (Kennie, 2014; Stevenson, 1994).

More advanced tools and algorithms emerged later to improve the extent of automation for line tracing. Eikvil *et al.* (1995) implemented an interactive tool for binary map image vectorization. With this tool, linear features are extracted and vectorized through tracing the contours at both sides of the lines. The course of the contours is stored as they are traced. The midpoint between the current two contour points is computed at short intervals, and stored to represent the line. Moreover, curvatures and corners are detected by checking whether the direction of the contour segments changes. For intersections, the type of intersection is determined by the user before subsequent line tracing. The tracer may determine the direction of the traced segments. Otherwise, the tracer stops and asks for user interaction.

Even until recently, interactive line tracing was still widely used and further developed, as it brings line tracing under human control and provides the ability to correct the data immediately if required. Yang *et al.* (2012) proposed a local adaptive segmentation method based on a sliding window to extract linear features, followed by a sequential line tracing process. Specifically, the starting point and an initial direction of the line to be traced are specified by the user, based on which a sliding window is created. The colour image in the sliding window is converted into a gray-scale image. Then, K-means clustering is used to segment the linear feature in a small neighborhood in the center of the sliding window. Subsequently, directional region growth is performed in the whole sliding window to segment the linear feature. A thinning operation is applied to the segmentation results. Sequential line tracing and vectorization are performed by moving the sliding window along the linear feature until an endpoint or an intersection is met. Automatic line tracing stops at an intersection point, and the next point is manually provided. The method may not work for tracing linear features symbolized by dashed lines.

In summary, due to the complexity of linear features (e.g., intersected road lines, dashed lines), interactive line tracing guided by a human operator is a practical and effective way of vectorization. Interactive line tracing is much more efficient and economic than manual line tracing, and more practical than automatic line tracing, as fully automatic extraction and vectorization of linear features is still far from being mature enough to be universally applicable. Usually, interactive line tracing works on binary raster maps. However, Howman and Woodsford (1977) point out that interactive line tracing works more effectively for single lines (e.g., road features symbolized with single lines, contour lines) than for double lines (e.g., road features symbolized with parallel lines).

### 3.2.2.1.2. Automatic line tracing

Single line tracing methods are developed to trace either the medial lines of linear features (roads, railways, and water areas) or the line segment array generated through a map vectorization process, thereby extracting road features from maps. Suzuki *et al.* (1987) traced and extracted the medial lines of road features by assuming that road features have the same width and small curvatures. Later, Suzuki and Yamada (1990) proposed an algorithm to trace the medial lines of road features by searching a line segment array that is obtained through vectorization. The array records the starting and ending points of each line segment, and line branches that contain the segment. Nonetheless, in these two studies, the vectorization process that generated the medial lines through the thinning operation may lead to distortion and lower the accuracy of the medial lines. Moreover, if road features have the same width as other linear features (e.g., railways, contours), these two methods may fail in distinguishing these features. A similar approach that traces and recognizes road features based on the line segment list is designed by Amin and Kasturi (1987). The line tracing algorithm is developed to order line segments constituted of connected pixels into lines based on geometrical constraints like segment length and orientation. Subsequently, road features are recognized through analyzing the line segment list. When a query is requested, the image processing routines are triggered. Then, operations such as skeletonization and line tracking are used to extract and display all the lines representing the queried roads. Similar methods to trace road lines are implemented in Alemany and Kasturi (1988), Kasturi and Alemany (1988), and Kasturi *et al.* (1989). Nevertheless, the application of the algorithms is limited to simple

maps, and several parameters have to be manually specified. Despite the effectiveness of the single line tracing approaches, the thinning operation that produces the medial lines of linear features often leads to line distortion, especially around road intersections (Chiang and Knoblock, 2008). More importantly, the single line tracing methods usually neglect the parallel characteristic of road lines, which is the most prominent characteristic of road features (Bin and Cheong, 1998; Liu, 2002), as the parallel characteristic has been lost in the vectorization process. In addition, most single line tracing methods did not take advantage of the colour information of raster maps.

Many other methods employ the parallel characteristic of road lines to extract road features from map images. Miyatake (1985) proposed a parallel line extraction method to extract road features. Specifically, the spacing width between the parallel lines is computed. The length of each connected component is set as the length of its surrounding rectangle. The pixels with labeled width and length values in a given range are regarded as the space between parallel lines. Next, the pixels adjacent to the space are extracted as parallel lines. Furthermore, road intersections are extracted through an expansion and contraction process. The medial lines of the extracted parallel lines are obtained with the tracing method proposed by Kakumoto *et al.* (1983). With this method, road features of different widths and road intersections can be extracted. Nevertheless, several parameters like the range of the spacing width between the parallel lines have to be set manually. Moreover, an interactive step to correct the road extraction results and to connect the disconnected lines is required. Nakajima *et al.* (1985) defined and used parallel vector tracers to extract road features. The tracers have two pairs of connecting vectors that move along the parallel edges of the road features. This algorithm succeeds not only in tracing the parallel edges, but also the intersections. Nonetheless, the starting points of the vector tracers and the initial length of each vector cannot be automatically determined. Liu (2002) introduced an algorithm named the “rolling ball” method, which can recognize and vectorize raster map features simultaneously. Specifically, parallel road curb lines are traced with the rolling ball method, in which a ball that always touches the two parallel road lines moves along the road branches. The center points of the rolling ball are the road centerline points. Namely, the centerlines of road branches are obtained with this method. Furthermore, an adaptive road inter-junction detector is developed to detect road intersections. Subsequently, road networks are constructed based on the road centerlines and the center points of the road intersections. Although this method generates promising road extraction results, it may fail in recognizing road branches symbolized with dashed lines, as the rolling ball always needs continuous road lines. Later, Chiang *et al.* (2005) proposed an approach that can automatically distinguish single- or double-line road features by varying the road width from 0 to 10 pixels for parallel pattern tracing. In this process, the road width can be automatically obtained for double-line road features. The foreground pixels that do not have the parallel properties for the given road width are removed. Moreover, road intersection points are detected by using the interest operator (Chen *et al.*, 2004) and filtered with the criterion of a connectivity (the number of lines intersecting at an intersection) of more than two. Nonetheless, there is distortion around each road intersection due to the morphological operation. Thus, this approach was improved by Chiang and Knoblock (2008) and Chiang *et al.* (2009). In Chiang and Knoblock (2008), road orientations are computed by tracing the road medial lines. Road intersection templates are then generated based on the road orientations to refine the positions of the road intersections. In Chiang *et al.* (2009), LTM (Chen *et al.*, 2008) is utilized to enhance the accuracy of the position, connectivity, and orientation of the extracted intersections. Specifically, a double-line road intersection template or a single-line template is constructed for each extracted road intersection. Then, the position of the intersection point is adjusted with this template based on LTM. Although the approaches achieve promising results in various maps, it seems that they may not work for road features symbolized with dashed lines, as the properties employed in parallel pattern tracing did not cover the condition of dashed lines.

The studies reviewed above fully employ the parallel characteristic by tracing the parallel road lines and can obtain the width of road features. Nonetheless, many of the methods may not work for road features symbolized with dashed lines. Apart from this, most of them did not employ the colour information of raster maps. These are the possible reasons why the parallel line tracing methods have rarely been developed and applied to road extraction in recent years.

### 3.2.2.2. Parallel characteristic extraction

Some other studies take advantage of the parallel characteristic of road lines to extract road features, although they do not directly trace the parallel lines, as the parallel tracers may stop where other features like characters cut off the road lines (Nagao *et al.*, 1988). To alleviate this problem, Nagao *et al.* (1990) proposed a skip-scan method to automatically extract the medial lines of parallel line road features. Specifically, a map is scanned by horizontal and vertical scan lines of constant intervals, respectively, so all the intersection points of a scan line and parallel lines are obtained. The middle points of these intersection points are connected to form the medial lines of road features. Experiments show that more than 90% of road features can be successfully extracted. Nevertheless, the width of road features has to be known before processing, and the location precision of the extracted medial lines is dependent on the interval of the scan lines. Moreover, it is not accurate to simply use straight lines to connect the broken road lines resulting from the elimination of characters.

Bin and Cheong (1998) developed a system to extract and generate road networks from urban maps. In this system, parallel lines are obtained through parallel grouping. Subsequently, road medial lines are extracted based on the parallel lines and topologically connected to construct the road networks. Despite the successful extraction and generation of road networks, it seems that the method did not automatically remove the medial lines resulting from either the parallel lines of building edges, or the building edges parallel with the road lines. Nishijima and Watanabe (1998) reported a study that applies the generation and verification paradigm of hypotheses to road extraction from urban maps. In this method, road networks are extracted and constructed by searching out pairs of successive parallel line segments. Similarly, in Watanabe and Oshitani (2001), the inferred road fragments are verified by searching pairs of parallel line segments. In the study by Dhar and Chanda (2006), road features represented by parallel lines are extracted by applying the Hough transform. This technique is further elaborated in the next section.

### 3.2.2.3. Hough transform

The Hough transform can be used to detect lines in images by detecting intersection points between the sinusoidal curves in the Hough space (Ballard, 1981). In the study demonstrated by Dhar and Chanda (2006), the Hough transform is used in the pre-processing step to compute the orientations of lines obtained from the thinning operation, which contributes to joining the broken lines. Apart from this, the Hough transform is applied to distinguish different road classes represented by a solid single line, two parallel solid lines, and two parallel dashed lines by finding two maxima at the same angle, but at slightly different positions regarding the distance to the image center.

Chiang and Knoblock (2009a) applied the Hough transform to identify Hough lines from the colour layers segmented from the user label image through CIS. The user label image is a small rectangle labeled by the user, which covers a road intersection or a road segment on the input map image. Then, the colours in the colour layers where the average distance between the detected Hough lines to the image center is within a threshold are selected as road colours. Nevertheless, a limitation of the Hough transform in this method is that its effectiveness relies on the number of detected Hough lines. If a colour layer has only a small number of road pixels, the corresponding colour will not be selected as road colour. Thus, an alternative technique has to be used to extract road pixels from the colour layer. The method is also applied in Chiang and Knoblock (2009b) and Chiang and Knoblock (2013).

### 3.2.3. Image filtering-based road extraction

#### 3.2.3.1. Conventional filters

An image filter or kernel is a small matrix used to carry out a convolution operation on an image in order to change its appearance. In this way, different effects, such as sharpening, blurring, embossing, edge detection, and so forth can be obtained. Conventional image filters can be used to remove noise. Liu (2002) applied median filtering to reduce map image noise in the pre-processing step. In the study by Chiang *et al.* (2011), noise in the road layer identified based on a user label and CIS is removed by employing proper image processing filters and generating parameter sets identified through incorporating user-provided “noise samples.”

Conventional filters play an important role in CIS. Yin and Huang (2001) employed the median filter to smooth the segmented images representing different classes of roads. The images are segmented through a gray-scale histogram. Chiang and Knoblock (2009a, 2009b) applied the mean-shift technique to reduce the number of colours in the input image. The mean-shift filter merges two colours into one if their distance in the RGB colour space and the spatial distance of the corresponding pixels in the image space are both within the preset thresholds. Considering that a Gaussian or median filter may result in a substantial loss of information about the position of the roads and their edges, Callier and Saito (2011) selected mean-shift to reduce the number of colours and noise in the input map image. Similarly, in Chiang and Knoblock (2013), mean-shift is used to reduce noise and the number of colours but to preserve the edges of map features (e.g., road lines). Mean-cut is then applied to further reduce the number of colours to 1,024 at most. The colour cluster boxes in the HSL space are continuously divided at the median until the total number of boxes is smaller than the desired number of colours. The colours in the same colour box are then represented by their median colour.

Conventional filters can also be used as auxiliary techniques in machine learning methods to extract road features. In the study reported by Saeedimoghaddam and Stepinski (2020), to enlarge the training dataset, a Gaussian filter is used to provide tiles with low graphical qualities for data augmentation. To analyze the effect of sharpness or blurriness of the input map image to the prediction results, a Laplacian filter is employed to compute the minimal blurriness.

#### 3.2.3.2. Morphological operations

A morphological operation is conceptually defined by moving a window (e.g., a shifting window of  $3 \times 3$  pixels) over the image, in such a way that it is eventually centered over every image pixel where a local logical operation is performed (Bovik, 2009). The most commonly used morphological operations in image processing tasks are binary operations, to which the input is a binary image. When the image is “scanned” by the shifting window, these morphological operations generate binary results based on the hit-or-miss transform. Eventually, the morphological operation creates a new binary image. Fundamental binary operators include erosion and dilation. The erosion operation removes small-scale details from a binary image but simultaneously reduces the size of regions of interest, too. The dilation operation has the opposite effect to erosion. It adds pixels to both the inner and outer boundaries of regions. Many morphological operations are represented as combinations of the erosion and dilation operators, such as opening and closing. The opening operation can open up a gap between objects connected by a thin bridge of pixels. The closing operation fills holes in the regions while keeping the initial region sizes. Especially the thinning operation is useful for producing the skeleton of a group of foreground pixels, and thus is often applied for extracting the centerlines of road areas (Chiang *et al.*, 2014).

Morphological operations are frequently applied in road feature extraction from map images, in order to get the road centerlines (Ahn *et al.*, 1997; Chiang *et al.*, 2009; Chiang and Knoblock, 2009a; Chiang and Knoblock, 2009b; Chiang and Knoblock, 2013; Itonaga *et al.*, 2003), to reduce noise (Ahn *et al.*, 1997; Chiang and Knoblock, 2013; Linton, 2009; Liu, 2002), to reconnect broken road lines (Chiang and Knoblock, 2009a; Chiang and Knoblock, 2009b; Chiang and Knoblock, 2013; Chiang *et al.*, 2009; Dhar and Chanda, 2006), and to refine the extracted road areas or road intersections (Ahn *et al.*, 1997; Chiang and Knoblock, 2013). In these methods, morphological operations are used as auxiliary techniques.

The studies that employ a morphological operation as a main technique to extract road features were initiated by Yamada *et al.* (1991). They proposed a concept of multi-angled parallelism (MAP), which unifies the two concepts of non-isotropic neighbors for feature orientation and directional elements. Each pixel of the map image is composed of multiple directional elements (e.g., 8 or 16 directions). The authors then define a set of MAP operations based on conventional erosion and dilation to extract linear features. The operations are performed upon multiple directional elements of the pixels. Specially, a directional erosional operation is performed to extract the road lines. Road areas are extracted using fan-shaped dilation operations, as road areas can be regarded as an overlapped region of expansion from two parallel lines. Nevertheless, the extracted linear features are separated at the curved sections, as in this method “line” was defined as a straight line segment. Thus, Yamada *et al.* (1990) solved this problem by connecting the line segments of long linear features using a set of directional dilation operations. Moreover, the broken parts in the long linear features are restored by using dilation operations. Road features are extracted using the directional dilation of facing directions as they consist of parallel lines. Nonetheless, some roads in the downtown areas are wrongly extracted as hatched regions (e.g., buildings). Therefore, Yamada *et al.* (1993) improved this method to solve the problem by using directional dilation to restrict the directional operation.

MAP may detect all parts of linear features at the expense of misclassifying segments of characters as linear features (Pezeshk and Tutwiler, 2010). Therefore, Pezeshk and Tutwiler (2011) demonstrated an approach to solve this problem. Specifically, they defined four primary directions instead of eight directions, as the strict directionality imposed by MAP results in over-fragmentation of lines into many short segments (Pezeshk and Tutwiler, 2010). Furthermore, line segment pixels that have neighbors in at least two of the adjacent primary directions are selected as seed pixels. These pixels are located in areas where the local direction of a linear feature is changing. The whole linear feature is obtained through seed growth and linking the line segments. The length criteria are used to separate linear features from character segments. Experiments show that the approach generates promising road extraction results from U.S. Geological Survey (USGS) maps. Based on this approach, a system for automatic extraction of various map features and recognition of the text content from scanned topographic maps is developed by Pezeshk (2011). Nonetheless, some straight segments of large characters were still misclassified as linear features.

Additionally, Itonaga *et al.* (2003) developed an approach for automatic extraction of road networks from urban planning maps. First, road areas are recognized by stochastic relaxation based on the geometrical properties of the areas. Then, a thinning operation is applied to the recognized areas to extract road centerlines. But the thinning operation results in geometrical distortion in the extracted centerlines, especially around the line intersections. To solve this problem, the position of the intersection point is updated based on the angle of a short line segment connected to the intersection area. Next, piecewise linear approximation is applied based on the corrected road centerlines, and the road network is constructed. Nonetheless, this method may not work for areas where road features are intersected with overpasses, labels, and so forth. Moreover, the quality of the extracted road networks depends on the setting of parameter values, which may make the method map-specific.

In summary, morphological operations are commonly used to modify image contents to facilitate the extraction of linear features. Usually, CIS is carried out on the map image to obtain a set of binary images representing individual map layers before the application of binary morphological operations. Nonetheless, they have some drawbacks. First, morphological operations (e.g., the thinning operation)

may distort the geometry of the original linear features, especially around line intersections. Thus, alternative techniques need to be developed to correct the geometric distortion (Chiang *et al.*, 2014). Moreover, some parameter values (e.g., the operation iteration, the length of line segments) usually have to be determined manually based on the attributes of the maps or through experiments. This makes the morphological methods map-specific. Additionally, morphological operations may work less well for extracting road lines symbolized by dashed lines.

### 3.2.3.3. Convolutional neural network

Machine learning methods provide an effective way to extract features from raster map images because of their excellent performance in classification. Some early machine learning methods are developed to obtain a set of feature values based on the training images to facilitate the extraction of road features. The road extraction method proposed by Yin and Huang (2001) computed the geometrical feature values of the map title box and the legend index table (e.g., the ratio of the longer side to the shorter side of the legend index table) from the training map images. Recently, CNNs, especially deep CNNs, have exerted their superiority in automated feature extraction from images. Deep CNNs contain tens or hundreds of successive layers that gradually extract complex features from an input image, and then predict the probability with which a certain area (e.g., whole image, single pixel, rectangles within an image) depicts a certain class. A deep CNN architecture usually consists of convolutional layers, different activation functions (e.g., rectified linear units), and pooling layers (e.g., max or average pooling) (Saeedimoghaddam and Stepinski, 2020).

One of the advantages of CNNs lies in their generality compared with other machine learning models for image feature recognition like SVM. CNNs have the capability to recognize different types of map features, or the same type of features represented by different symbols (e.g., different classes of roads represented by different symbols) (Duan *et al.*, 2017). Yet, CNNs require a large amount of training data to perform sufficiently well. Thus, it is laborious and time-consuming to manually create the training data. To alleviate this problem, Duan *et al.* (2017) proposed an algorithm that automatically generates training data to facilitate the subsequent extraction of railroad features from USGS historical maps using CNNs. The algorithm automatically and accurately aligns contemporary railroad vector data with the corresponding railroad features on the maps. Later, Duan *et al.* (2020) presented another automatic vector-to-raster alignment algorithm to generate training data for the extraction of road features from USGS historical maps. This algorithm models the alignment problem using the reinforcement learning framework to precisely annotate the locations of road features on the maps. Nonetheless, the algorithm did not move the adjacent vector road segments as a group, resulting in losing geometric and topological information of intersections, or distorting the road orientation. To compare the impact of CNN architectures on feature extraction accuracy from raster maps, Chiang *et al.* (2020) presented a set of experiments for railroad extraction from USGS historical maps. Although the railroad features are successfully recognized from the map images, the experiment results show a limitation of CNNs, that is, the convolutional and pooling layers included in CNNs make it difficult to recover the detailed spatial locations of map features, especially the locations of small features.

A study that employs deep CNNs for road intersection extraction is reported by Saeedimoghaddam and Stepinski (2020). They adopted the faster region-based deep convolutional neural network (RCNN) framework to extract road intersections from USGS historical maps. Specifically, the faster RCNN framework first uses a deep CNN to extract the feature maps of the map image, followed by an implementation of a region proposal network (RPN) in order to select excellent candidates from the feature maps. Then, the selected candidates are fed into two fully connected layers to compute the probability of the candidates being a road intersection and to refine their bounding boxes. In this study, the authors used a pre-trained deep CNN to reduce the training time. Moreover, as data size is the key factor in deep CNN performance, the authors enlarged the training dataset using data augmentation

techniques. Experimental results demonstrate that road intersections represented as both single lines and double lines can be successfully extracted. Nonetheless, road branches cannot be extracted.

Despite the rapid development and superiority in feature recognition and extraction of CNNs, there exist up to now only limited numbers of studies that apply CNNs to road feature extraction from raster maps. It should be explored how CNNs can be fully applied to the problem of road extraction, including how the parameters of the CNN architecture affect the road extraction results, how the characteristics of the map images (e.g., colour diversity, blurriness) impact the road extraction accuracy (Saeedimoghaddam and Stepinski, 2020), and how to use CNNs for generating large amounts of high-quality training data.

### 3.2.4. CIS-based road extraction

CIS separates thematic map layers based on homogeneous colour information, as thematic layers in maps, such as road networks, hydrography, vegetation, and so forth, are normally represented by a distinct colour (Leyk and Boesch, 2010). CIS is of critical importance since the outcome directly determines the image processing methods to be applied in all subsequent stages of map feature extraction (Chiang *et al.*, 2011; Leyk, 2009).

#### 3.2.4.1. Histogram technique

The histogram technique can be used in automatic CIS, aimed at separating the different colour layers in a map without user intervention (Chiang *et al.*, 2014). In a study by Ebi *et al.* (1994), the scanned raster maps are segmented into colour layers before the recognition of map features. First, the RGB data of the map image are transformed into the u'v' chromaticity plane ( $L^*u^*v^*$  colour space), and the u'v' histogram is generated. The peaks and ridges are detected in the histogram, from which the colour cluster centers are derived. The map image is then segmented using the cluster centers based on chromaticity and lightness criteria. Subsequently, region growth techniques are applied to correct the defects caused by overprinting with other layers (e.g., black tree symbols printed over a green forest region). After obtaining the colour-homogeneous layers, the geometric properties of map features are used to detect whether a layer contains mainly region or line structures. The line-based layers (e.g., the road layer) are thinned and vectorized to produce the medial line. The methods succeed in distinguishing road features from contour lines, as road features are in the black layer, while contour lines are in the brown layer. Nevertheless, if road features and contour lines have similar colour and are segmented in the same layer in the CIS process, the methods may fail in distinguishing them.

Another approach that utilizes histogram-based CIS is demonstrated by Yin and Huang (2001). In this approach, based on the histogram of the gray-level map image, the gray-level distributions of map features are analyzed using the multilevel thresholding technique, so different classes of roads (e.g., national highways, county roads) are segmented into separate layers. The extracted road features are vectorized, followed by a post-processing to restore the broken road lines by analyzing their slopes and the endpoint locations.

Histogram analysis is used to separate the foreground from the background pixels in the input map image (Chiang *et al.*, 2005; Chiang *et al.*, 2009; Chiang and Knoblock, 2008). The authors analyze the shape of the grayscale histogram. First, the largest luminosity cluster in the histogram is identified as the background cluster. Then, other clusters are classified as either background clusters or foreground clusters based on the number of pixels in the clusters.

A colour histogram segmentation approach is reported by Henderson and Linton (2009) to extract road pixels from USGS maps. First, the colour usage information (e.g., the number of pixels of the same colour) is retrieved from the map legend, based on which different colour layers are segmented.

Furthermore, geometric properties including spatial proximity, continuity, and closure are employed in a tensor voting method to find roads and intersections in the segmented layers. Nevertheless, this approach requires specific knowledge about the use of colours in the maps (e.g., the RGB value of each colour) as the basis for a colour histogram segmentation. Moreover, the parameter used in tensor voting depends on the size of road features, which seems too map-specific. The approach only analyzed a set of  $200 \times 200$  sub-images from USGS raster maps. Similar approaches are presented by Linton (2009) and Henderson *et al.* (2009). Another similar colour histogram-based map image segmentation approach is demonstrated by Henderson (2014). In order to generate an initial feature extraction result, the histogram model of the feature is created as a set of sample histograms representative of the feature class from the map legend. This is a pre-processing step for the subsequent analysis by tensor voting. More importantly, this study made a detailed comparison between the techniques of histogram-based classification as well as the techniques for extracting road features from the road curve map produced by the tensor voting process. The curve map gives the likelihood of the presence of a road curve passing through each pixel. The study also discussed how to approximate ideal parameter values for tensor voting.

Another study that employs histogram-based CIS is reported by Callier and Saito (2011) to extract road features from tourism maps. First, the mean-shift is used to reduce the total number of colours in the raster map. The method then searches for possible lines for each pixel in the  $15 \times 15$  shifting window centered at the pixel. If a line is found to have almost the same colour, then it is considered as a possible linear feature. Pixels with high probability of being a road are selected as seed points and seeded region growing is applied to find other possible road pixels. To extract complete road features, two three-dimensional histograms of the colours of the detected road pixels and the background pixels are created. The colours corresponding to the main peaks in the road pixel histogram are selected as road colour. Based on this colour, the undetected road pixels in the previous steps are retrieved. But it can be challenging to find reliable parameters for determining the same colour and selecting pixels with high probability of being a road, and the parameter values vary between different types of maps.

The histogram technique is frequently used in automatic CIS that contributes to the successful extraction of road features in subsequent image processing and feature recognition steps, but assumes high levels of homogeneity within colour layers (Chiang *et al.*, 2014). Automatic CIS techniques may be map-specific, as it can be challenging to choose the number of colour clusters and the values of other parameters. Moreover, the histogram technique usually requires combination with other image processing techniques.

### 3.2.4.2. K-means

K-means is an unsupervised machine learning method that is aimed at classifying the dataset into  $k$  pre-defined clusters. K-means repeatedly divides the data into  $k$  clusters according to a certain distance function until an optimization function reaches convergence.

In the study by Dhar and Chanda (2006), K-means is applied to Indian survey maps to separate different colour layers (e.g., red layer for man-made structures like roads). Specifically, K-means clusters the RGB colours of the image pixels. Nonetheless, in the clustering process,  $k$  is preset to five, which can be map-specific. Henderson *et al.* (2009) reported a study that employs K-means to segment semantic classes (e.g., roads) in raster maps. In the method proposed by Chiang and Knoblock (2009a) and Chiang and Knoblock (2009b), K-means is applied to the map image processed by mean-shift to further reduce the number of colours to not larger than the preset  $k$  value. Similarly, Chiang and Knoblock (2013) applied K-means to the image processed by mean-shift and median-cut to further merge similar colours. Yang *et al.* (2012) applied K-means clustering to a small neighborhood ( $5 \times 5$ ) to segment foreground pixels from background pixels. To facilitate the alignment of contemporary vector data to the features on historical maps, Duan *et al.* (2020) used K-means to group the pixel colours into clusters to detect the dominant pixel colours overlapping with vector segments. The clusters are used to formulate the

reward function in the reinforcement learning framework. If a cluster center is not within the colour range of the target map feature (e.g., road, water body), the reward for the segment is 0; otherwise, it is 1.

K-means is a simple and unsupervised technique in CIS. Sometimes, it requires other techniques as pre-processing steps, like median-cut to reduce the runtime (Chiang and Knoblock, 2013) and image enhancement to reduce colour variations (Dhar and Chanda, 2006). Nonetheless, limitations of K-means include that it may be difficult to forecast the number of clusters, namely the value of  $k$ , and that the clustering result is highly influenced by the original input (e.g., the value of  $k$ ). With a small  $k$  value, K-means may merge different semantic features (e.g., roads and text), as it considers only the colour space (Chiang and Knoblock, 2009b).

### 3.2.4.3. Other clustering techniques

Other techniques can also be used in CIS, and CIS can be developed as an interactive procedure, because it requires user input to indicate the colours of road pixels. Ahn *et al.* (1997) demonstrated a road extraction method based on interactive CIS for Korean topographical maps. The user needs to manually specify points that contain the colours of road features, and the center of the colour cluster is calculated. The pixels that have the shortest distance to the center in the colour space are segmented from the map image into a separate layer, which is a binary image. For example, contour lines and roads are coloured in red. Thus, they are segmented into one layer. Next, opening and thinning operations are applied to remove contour lines and obtain centerlines of road features, respectively. Vectorized road data are obtained by tracing the centerlines. Nevertheless, this method did not correct the distortion around road intersections resulting from the thinning operation.

To alleviate the parameterization problem in CIS methods and the colour variation problem in USGS maps, Leyk (2009) proposed a two-stage colour sampling approach. The first stage is implemented for the derivation of the colour value centroid in the colour layer based on colour value sampling. With these colour centroids, homogeneous regions are extracted based on their minimum distance to the centroid in the colour space. The black layer contains thematic objects such as road infrastructure, buildings, administrative boundaries, and characters. But the colour centroids for black areas can be very different from those of black linear features. For example, buildings and road lines both belong to the black layer but appear in different colour tones due to colour bleaching and aliasing of the historical maps. Thus, the second stage is aimed at classifying the parts that suffer from colour deviation by resampling of colour values for the adjustment of the colour centroids obtained in the first stage. Then, a post-processing step is implemented to generate a cartographic representation of road features.

Chiang and Knoblock (2009a) developed a supervised method that requires user input for extracting road pixels from raster map images. After the number of colours in the input image is reduced, the user needs to provide a user label for each road colour in the map. Subsequently, each user label is processed by employing the Hough transform and a template matching technique, so that a colour filter with all identified road colours is generated. Next, all the road pixels are extracted using the identified road colours. Despite the successful extraction of road pixels, the method is considered incomplete, as the geometry of road features is not extracted and vectorized. Thus, the method is improved in Chiang and Knoblock (2009b). In the improved method, morphological operations are applied to generate road centerlines based on the extracted road pixels. A problem of using the morphological operators is that the thinning operator usually distorts the lines near the intersections. For correcting the distortion and generating accurate road vector data, the authors detect the intersections of the thinned lines and trace the lines outside the distorted areas to generate accurate road orientations and intersection positions. Similarly, in the method reported by Chiang *et al.* (2011), a user provides a “road sample” through labeling a sample area centered on a road line, so that the features in the same colour as the labeled road lines are automatically recognized. The recognized results are refined by employing user labels that provide samples of road and noise pixels to remove the non-road pixels through image processing filters.

Subsequently, the refined road features are vectorized. Nonetheless, the refined road features still contain some undesired pixels (e.g., grid lines, characters) and broken road lines, which indicates that a further refinement step or manual post-processing is needed. Moreover, the quality of the CIS results relies on the parameter value of the number of desired map layers.

Chiang and Knoblock (2013) demonstrated a general road vectorization approach by integrating and improving the approaches reported in their earlier work (Chiang *et al.*, 2009; Chiang and Knoblock, 2009a; Chiang and Knoblock, 2009b). Specifically, road pixels are extracted using interactive CIS (Chiang and Knoblock, 2009a). The authors improved the time complexity of the parallel pattern tracing algorithm reported by Chiang *et al.* (2009) and developed a single-pass parallel pattern tracing algorithm to detect the road width and road format. Next, morphological operations are used to generate road centerlines based on the detected road width and road format. As the thinning operator usually distorts the lines near the intersections, the method then traces the road lines outside the distorted areas (Chiang and Knoblock, 2009b). The locations of the road intersections are updated using the traced lines. Subsequently, to generate the road vector data, the road centerlines are traced based on the accurate positions of the road intersections (Chiang and Knoblock, 2009b). Nevertheless, one limitation of the method is that the vectorization process relies on the width of the majority of road features. As a result, some small road branches are eliminated by mistake, as their width is smaller than the width of the majority of roads. Moreover, a post-processing step is required to reconnect the broken road lines.

The clustering techniques in CIS generate a colour filter for the subsequent map feature extraction. Usually, other image processing techniques (e.g., Hough transform, morphological operations) are required in the subsequent steps to refine the results of CIS and extract the geometry of target features.

### 3.2.4.4. Support vector machine

SVM is a machine learning method which is used for the separation of map layers based on colour information or the separation of foreground and background pixels. SVM was first proposed for classification and regression analysis. SVM solves classification problems by finding an optimal hyperplane for linearly separable data, and is extended to non-linearly separable data by transforming the original data to a higher-dimensional space with kernel functions (Ben-Hur *et al.*, 2001).

Chiang and Knoblock (2006) reported a study that employs SVM in combination with discrete cosine transformation (DCT) coefficients to extract road pixels from raster maps. First, foreground pixels are separated from the background through generating the DCT coefficients for pixels on the input raster map. Based on the property of consistent colour in the background, pixels with low DCT coefficients are classified as background. The second stage is to classify road and character pixels among the foreground pixels. As characters are generally more complex than road lines, the DCT coefficients of character pixels are higher than those of road lines. Thus, the authors generate the DCT coefficients for each foreground pixel and send them to the SVM for classification. The authors use two street maps from Google Maps and one from ViaMichelin as road training data and manually remove the characters from them. The misclassified pixels are corrected by performing connected component analysis. Nonetheless, as the foreground pixels include other map features than road lines and characters, the approach still needs refinement.

In summary, CIS plays a key role as a pre-processing step in the whole workflow of road extraction. However, CIS may fail in separating thematic layers if different map features share the same colour. For example, brown pixels in the USGS topographic maps are used for both the contour lines and roads. Thus, many of the road extraction results include contour lines (Chiang and Knoblock, 2009a). Moreover, semi-automatic CIS requires the user to provide enough user labels to cover each road colour in the raster map. Usually, the user label has to cover a road segment or intersection and should be located at the center of, or just a few pixels from the center of a road segment or intersection (Chiang *et al.*, 2011; Chiang and Knoblock, 2009a; Chiang and Knoblock, 2009b; Chiang and Knoblock, 2013).

### 3.3. Road feature extraction from overhead imagery

In this section, road extraction methods from overhead imagery which can possibly benefit road extraction from raster maps but have not yet been applied to raster maps are reviewed. As a reference for enhancing road extraction performance from raster maps, this section mainly focuses on the new methods and achievements from 2014, as since 2014 special attention has been shifted to deep learning in the remote sensing community (Ma *et al.*, 2019). Furthermore, as criteria to evaluate road extraction results, evaluation methods are also reviewed. Compared with the monotonous evaluation metrics for road extraction from raster maps, there are various metrics for the overhead imagery, which will inspire researchers to possibly apply and tailor the metrics to raster maps.

Roads in overhead imagery and in raster maps are long, extended slender areas (Chen *et al.*, 2017; Sun *et al.*, 2019; Wang *et al.*, 2016; Zhou *et al.*, 2018). Ideally, roads should be extracted as continuous, connected, and intersected long lines that can form a network. However, it is usually challenging to preserve the continuity and topology of road features, as in the overhead imagery, roads are occluded by trees and shadows (Wang *et al.*, 2016), and in the raster maps, roads are interrupted by labels and other features (e.g., water bodies, railroads). Thus, different strategies are proposed and applied to address this issue.

#### 3.3.1. Machine learning architectures

Some studies adapt machine learning architectures in order to enhance their abilities to solve the discontinuity issue of road extraction. Zhou *et al.* (2018) propose an encoder–decoder neural network structure, named D-LinkNet, which inserts dilated convolution layers between the encoder and the decoder to address the challenge of road extraction from high-resolution satellite imagery by using dilated convolution to enlarge the receptive field of filters. The dilated convolution layers are stacked both in cascade mode and parallel mode in order to take advantage of and combine multi-resolution features. ResNet34 (He *et al.*, 2016) pre-trained on the ImageNet (Deng *et al.*, 2009) dataset is deployed as the encoder, as it is found that transfer learning can accelerate the convergence of the network and make it perform better. D-LinkNet achieved top performance in the CVPR DeepGlobe 2018 Road Extraction Challenge. Nonetheless, the discontinuity issue of the extracted road features is not completely addressed (Abdollahi *et al.*, 2020).

Inspired by the U-net (Ronneberger *et al.*, 2015) and atrous spatial pyramid pooling (ASPP) (Chen *et al.*, 2017) approach, He *et al.* (2019) integrated ASPP into U-net in order to grasp multi-scale road characteristics such as local corners, textures, macroscopic lines, and global network structures, as atrous convolution is capable of adjusting the receptive field of the filter. Specifically, the ASPP module used in this article consists of one  $1 \times 1$  convolution and three parallel  $3 \times 3$  convolutions with atrous rates of 6, 12, and 18, respectively, in combination with an image-level pooling layer. Placed after the bottleneck of the encoder–decoder network, ASPP is applied to the feature map produced by the encoder, and the resulting feature map is fed into the decoder. Nonetheless, the pooling layer may reduce the resolution of center feature maps and drop spatial information.

Tao *et al.* (2019) designed a spatial information inference structure (SIIS), enabling them to extract and transmit not only local road characteristics, but global and contextual road information in four directions. The SIIS is inserted after the bottleneck of the DeepLabV3+ network (Chen *et al.*, 2018). Specifically, the input of SIIS is a set of feature maps produced by the encoder. The feature map set is split into chunks along two dimensions, and the chunks are fed into a 3D convolutional recurrent neural network (Conv3d-RNN) one by one. The Conv3d-RNN is developed by replacing all the matrix operations in the traditional RNN unit with 3D convolution. Despite the effectiveness and robustness of the SIIS-Net, it may fail in extracting some very narrow country roads.

### 3.3.2. Alternative loss functions

Another way to preserve the continuity and topology of roads is to adapt and improve the loss function, as the normal loss functions, such as cross-entropy, assign equal weights to each pixel, thus ignoring the spatial and topological information when evaluating the similarity between the predictions and the ground truth (He *et al.*, 2019; Mosinska *et al.*, 2018).

Wei *et al.* (2017) propose a road structure-based loss function that incorporates the geometric information of road features in cross-entropy loss through imposing a large penalty of loss on the pixels close to road regions while imposing a small penalty on pixels far from road regions.

Mosinska *et al.* (2018) propose the topology loss as a supplementary term to the binary cross entropy (BCE) loss, which is aware of the higher-order topological characteristics (e.g., connectivity, continuity) of linear features. The feature maps obtained from several layers of a pre-trained VGG19 network (Simonyan and Zisserman, 2014) are used as a description of the higher-order characteristics. The topology loss tries to minimize the differences between the VGG19 description of the ground truth and the corresponding prediction. Experiments show that the prediction performance is increased by using this new loss function without having to change the network architecture.

He *et al.* (2019) proposed the structural similarity loss that evaluates the similarity between two images by comparing their luminance, contrast, and structure. The luminance is compared based on the mean intensity of pixel values, the contrast compared by using the standard deviation, and the structure information compared by computing the correlation (inner product) of the normalized pixel values. The statistics of the local structural similarity are calculated by using a square sliding window. The structural similarity is added as a term to the BCE loss and the obtained loss is used to train the network. Nonetheless, the hand-designed metrics and the topology loss computed based on the pre-trained VGG19 may be hard to generalize.

### 3.3.3. Data fusion

One common problem of machine learning methods lies in the fact that it is usually time-consuming and tedious to manually label and sample the training data. Thus, some studies fuse other data as the complement to the overhead imagery for training. Wang *et al.* (2015) used a vector-guided sampling strategy for generating training data for road extraction. Specifically, in the step of preparing the training data, cubic B-splines are employed to refine the vector road lines, so the road polylines visually match with the roads on the aerial imagery. Sample points, centered on which the image patches are to be extracted, are classified according to whether they locate in road areas and the angle of the road segment. A local image patch is extracted based on the location of a sample point and the angle information of the point. Then, the labeled image patches are fed into a DNN for road pattern recognition and a finite state machine (FSM) is used for tracking the roads on the imagery. Nonetheless, the tracking process is manually triggered by selecting an initial position and orientation for the tracker.

Sun *et al.* (2019) proposed fusing GPS data with aerial imagery for road extraction. Specifically, the GPS data are rendered as new input layers and fed together with the RGB channels of the imagery into the encoder-decoder network. To overcome over-fitting, a novel way of data augmentation is applied to the GPS data, including subsampling, resolution reducing, random perturbation, and data omitting. Instead of the conventional  $3 \times 3$  transpose convolution filters, 1D transpose convolution is used in the decoder, as the 1D filters are more aligned with road shapes, thus contributing to reducing gaps in the road extraction results.

Zhang *et al.* (2020) used GPS trajectories of floating car data as training data to extract roads from high-resolution remote sensing imagery, which avoids the tedious and time-consuming manual labeling process. Specifically, the GPS trajectory data are first rasterized and then denoised with morphological

operations. Next, the trajectory data are matched to the road in the remote sensing imagery in terms of resolution and road width. The results show that roads occluded by buildings or trees are extracted, verifying that the method is able to preserve the continuity of road features to some extent.

Nevertheless, machine learning methods still cannot completely tackle the challenges in road extraction. Highly accurate road extraction results are still not achieved. Issues like fuzzy boundaries as well as small and dispersed false positives still remain challenging, as CNNs mainly count on texture and spectral features, and the mixed pixels in road borders lead to misclassification. Moreover, many roads extend a far distance in the imagery, which demands high-level semantic information (e.g., multi-scale features) to preserve their completeness and continuity. Furthermore, pre-and post-processing operations are still necessary to achieve satisfactory results (Abdollahi *et al.*, 2020).

### 3.3.4. Shape descriptors

Shape descriptors can be used to describe accurately road-specific geometric properties (e.g., narrowness, the parallel characteristic of road edges), thereby effectively recognizing roads from the overhead imagery. Nonetheless, there are only limited numbers of studies that utilize shape descriptors to extract roads from raster maps. Thus, the methods that use shape descriptors for road extraction from overhead imagery are reviewed.

For many years, shape descriptors have played an important role in road extraction from overhead imagery (e.g., Mayer and Steger, 1998; Steger, 1998; Steger, 2000). Recently, shape descriptors have served as a supplementary technique in machine learning. Li *et al.* (2018) define two shape descriptors to automatically recognize roads and filter outliers from the results obtained through superpixel segmentation. Specifically, deviation of parallelism (DoP) is defined as the deviation of the width of a superpixel to describe the parallel characteristic of road edges, and narrow rate (NR) as the ratio of the length and width of the superpixel to describe the long and narrow characteristic of roads. Despite the effectiveness of this method, it may be difficult to obtain proper parameter values. Moreover, it may be challenging to precisely describe roads so as to quickly and accurately extract roads (Wang *et al.*, 2016). Nonetheless, usually raster maps are less complicated than remote sensing imagery. Therefore, shape descriptors can still play a role in road extraction from raster maps.

### 3.3.5. Evaluation metrics

Standard evaluation metrics for road segmentation results include recall, precision, quality, and so forth. The computation of the metric values is based on the number of correctly or wrongly segmented pixels, namely true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). Recall that the percentage of TP over all target pixels, namely TP plus FN, describes how completely roads are extracted. Precision, defined as the percentage of TP over all extracted pixels, namely TP plus FP, evaluates how correctly roads are detected. If the predicted road pixels as well as the ground truth are skeletonized or vectorized as road centerlines, completeness and correctness are more suitable than recall and precision, as they take into account the buffer of the centerlines and are therefore regarded as relaxed variants of recall and precision (Mosinska *et al.*, 2018). If the ground-truth centerline lies within a buffer of the predicted centerline, it is deemed TP, and FN otherwise. The lengths of TP and FN are used to compute completeness, which is  $TP/(TP + FN)$ . If the predicted centerline lies within a buffer of the ground-truth centerline, it is TP, or FP otherwise; and  $\text{correctness} = TP/(TP + FP)$  (Cardim *et al.* 2018; Wang *et al.*, 2015; Wegner *et al.*, 2013; Zhang *et al.*, 2020). Quality and F1 score, regarded as combinations of recall and precision or completeness and correctness, reflect the overall performance (He *et al.*, 2019; Li *et al.*, 2018; Mosinska *et al.*, 2018; Saeedimoghaddam and Stepinski, 2020; Tao *et al.*, 2019). Quality is estimated as  $TP/(TP + FP + FN)$ , and F1 score as  $2TP/(2TP + FP + FN)$ .

Accuracy is an intuitive metric that refers to the ratio of the number of correctly classified pixels to the number of all pixels (Wei *et al.*, 2017; Wulamu *et al.*, 2019). But if there are only a few target pixels on the image, the value of accuracy may not coincide with the effectiveness of the segmentation results. An alternative for accuracy is “intersection over union” (IoU), which, also known as the Jaccard index, refers to the intersection of the prediction and the ground truth divided by the union of the prediction and the ground truth (Sun *et al.*, 2019; Tao *et al.*, 2019; Wulamu *et al.*, 2019; Zhang *et al.*, 2020; Zhou *et al.*, 2018).

The above mentioned metrics may fail in evaluating the continuity and topology of the extracted road features. Thus, to evaluate the continuity, Tao *et al.* (2019) compute the number of road breaks, which represents the number of false fractures in the predictions compared with the ground truth. To evaluate topology, Mosinska *et al.* (2018) and Wegner *et al.* (2015) randomly and repeatedly sample pairs of connected points in the ground truth as well as in the predicted road network, and compare their path lengths. Incorrect gaps in the extracted network cause too long paths, or they disconnect the network into disjoint parts with no connection at all. Incorrect shortcuts result in too short paths. A tolerance parameter is pre-defined to account for geometric uncertainty. Paths with length difference smaller than the tolerance are regarded as correct. The point pairs are sampled until the percentages of these three error types have converged.

There exist similarities and differences between road extraction from overhead imagery and raster maps. A raster map is an electronic map image made up of pixels, which serves as a symbolic depiction of the geographical objects on the land surface. Raster maps can be made by scanning paper maps or by cartographic software. Overhead imagery refers to the images of the earth surface captured by imaging sensors. Features on raster maps are depicted by abstract symbols with various shapes, colours, and so forth, while features on overhead imagery are presented in their natural form. Texture, spectral, and (potentially) 3D information is captured in overhead imagery. Especially, roads are slender, long, and intersected areas in overhead imagery. Roads are recognized and distinguished from other geographical objects mainly by the shape and spectral information. Roads can be long, extended, and intersected areas or lines on raster maps, and are distinguished by symbolic and colour information. Roads can be occluded by trees, shadows, and interchanges on overhead imagery, while they can be interrupted by labels and point symbols (e.g., triangulation points). Line tracing works for road extraction from both data sources (Chiang and Knoblock, 2013; Wang *et al.*, 2015). Multi-scale spatial information will benefit both tasks, and the same evaluation metrics, as well as vectorization methods, can be shared by them.

### 3.4. Discussion

#### 3.4.1. An analysis of the development and progress of the road extraction methods of raster maps

The problem of road extraction from raster maps has gained much attention, and the relevant studies and methods have experienced steady development over the past 35 years. Up to 2020, 47 different road extraction methods based on raster maps have been identified. The total citation count of the 47 papers is 1,415 in Google Scholar. Figure 3.1 (a) and (b) present yearly paper counts and yearly citation counts, respectively. Figure 3.1 shows consistency between the paper counts and the citation counts. The paper count experienced two peaks in the late 1980s and around 2010, respectively. Accordingly, the citation count has two peaks in the 1990s and 2010s, as the citation count is lagging behind the paper count.

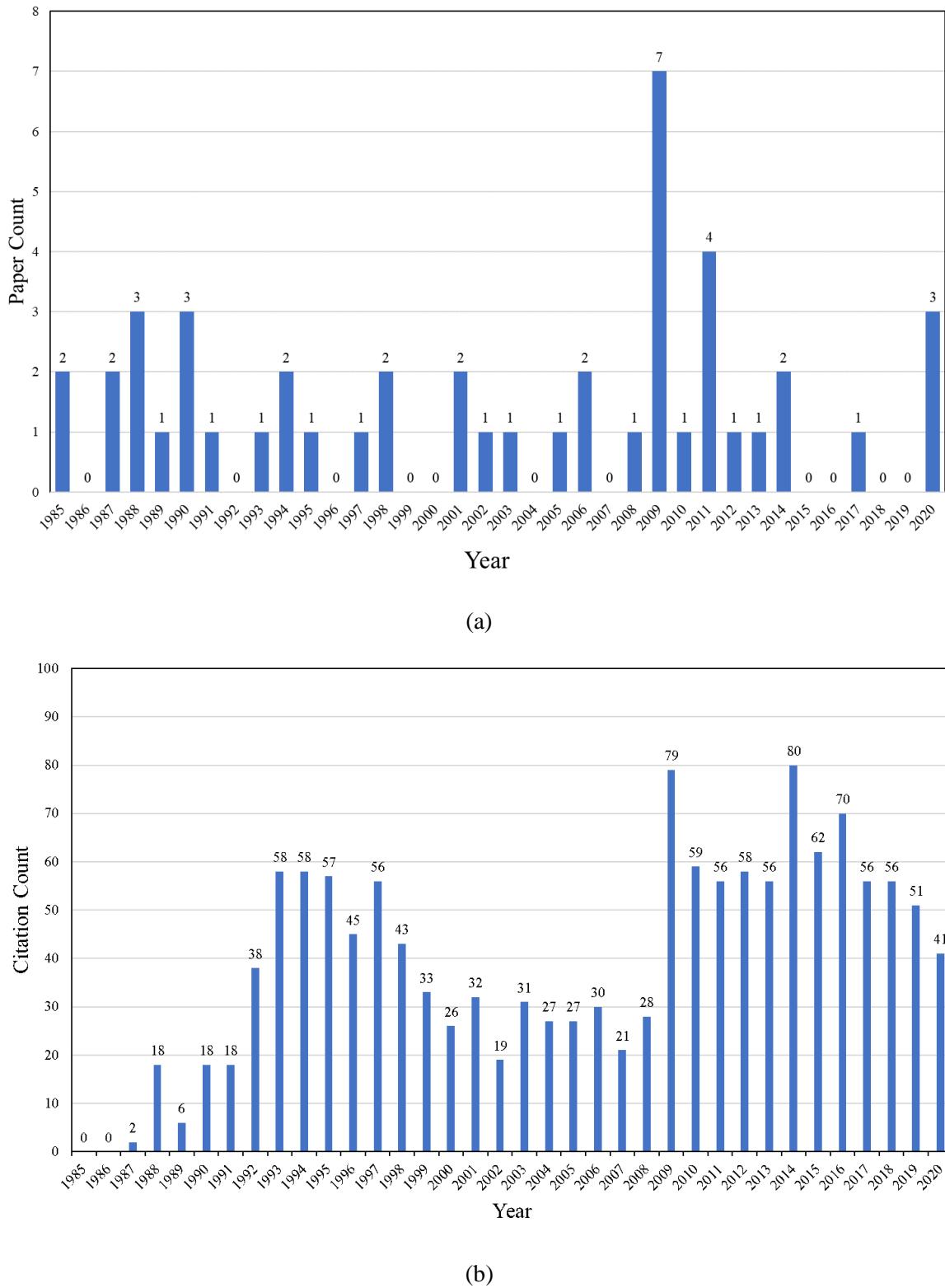


Figure 3.1. (a) Yearly counts of papers on road extraction from raster maps, and (b) citation counts of these papers. The search was conducted on September 27, 2020 using Google Scholar.

Interactive line tracing methods dominated the road extraction methods before the emergence of automatic methods. From the order of the techniques applied in the early methods, it is found that line tracing or parallel characteristic of road lines was directly applied without pre-processing like colour segmentation, as in early times raster maps were black and white or gray-scale images. For example, the

methods that only use interactive line tracing are clustered and coloured pink in Table 3.1. Later, automatic line tracing methods were intensively studied in the late 1980s and early 1990s. For example, the cluster of the methods that first applies line tracing and then morphological operations is coloured yellow in Table 3.1. In these methods, connected pixels are traced, so that line segments are obtained. Then, roads are recognized through an analysis of the line segment list. Especially, the authors implemented a road query function, in which the thinning operation is used to skeletonize the roads, so that the thinned line is displayed as the query result (Alemany and Kasturi, 1988; Amin and Kasturi, 1987; Kasturi and Alemany, 1988; Kasturi *et al.*, 1989). In recent years, line tracing methods were not frequently studied, probably due to the complexity of the line tracing process as well as the emergence and development of other automatic methods. Interestingly, the studies that use morphological operations to extract road features from raster maps emerged in the early 1990s, with the concept of MAP proposed by Yamada's research group. This method cluster is coloured orange in Table 3.1. Based on conventional erosion and dilation, the authors define a set of MAP operations to extract linear features (Yamada *et al.*, 1990; Yamada *et al.* 1991; Yamada *et al.* 1993). These methods were further improved by Pezeshk and Tutwiler in the early 2010s, mainly to tackle the problem of misclassifying character segments as linear features by reducing the number of primary directions of MAP as well as selecting pixels where the direction of a line changes as seed pixels. The whole linear feature is then obtained through seed growth and linking the line segments. Subsequently, the length criteria are used to filter linear features (Pezeshk, 2011; Pezeshk and Tutwiler, 2010; Pezeshk and Tutwiler, 2011). This cluster of methods is coloured purple in Table 3.1. On the one hand, the progress of the methods indicates that the previous studies provide directions and lay foundations for the later studies. On the other hand, the studies that mainly use morphological operations seem to be limited to MAP. Importantly, morphological operations are frequently used in road extraction because of their versatility. They can be used for noise removal in pre-processing, for reconnecting broken road lines, for skeletonizing the recognized road areas, and so forth. Conventional filters, like the median filter, were used for pre-processing the images in the early 2000s, as they work for noise removal and image smoothing. Later, CIS, especially the histogram technique and K-means, became more popular for pre-processing than conventional filters. For example, a set of CIS methods was presented by Henderson and Linton, in which different coloured layers are first separated based on the colour usage information retrieved from the map legend, and then geometric properties (spatial proximity, continuity, and closure) are used to detect roads and intersections in the separated layers (Henderson, 2014; Linton, 2009). This cluster is coloured bright green in Table 3.1. With the emergence and development of modern algorithms, various techniques are applied in one method to get promising road extraction results. For instance, the methods proposed by Chiang's group involve four different techniques, including the histogram technique as a pre-processing step to segment the foreground pixels from the background pixels, parallel pattern tracing to detect road pixels, morphological operations to reconnect adjacent road pixels, and skeletonizing the road areas (Chiang *et al.*, 2005, 2009; Chiang and Knoblock, 2008). This cluster of methods is coloured gray in Table 3.1. In recent years, machine learning has become a study hotspot and dominates the road extraction methods. Machine learning methods are used either for road feature extraction or automatically producing training data that facilitate subsequent road feature extraction. Usually, CNNs are applied without any pre-processing. CNNs stand out in these methods, owing to their superiority in feature recognition and extraction from images, as well as the wide and intensive attention paid to CNNs. In early times, there were often clusters of very similar road extraction methods. From 2010, there are rarely clusters of methods with the emergence of more modern techniques, indicating a trend of road extraction methods developing in a diverse way. Although machine learning has become a research hotspot, different techniques are applied, different inputs are required, and different goals (e.g., to extract road pixels, to generate road training data) are achieved.

Notably, more attention has been paid to historical maps since around 2009. Historical maps have become an important data source of the study on map interpretation and spatial feature extraction. However, the poor quality of historical maps and the demand for increasing feature extraction accuracy make it urgent to propose more advanced methods for recognizing road features and small objects (e.g., dashed line segments) (Chiang *et al.*, 2020). For instance, due to the spatial distortion inherent to historical maps, there usually exists a shift between the same road feature on the map sheets of different

years. How to automatically recognize the same road with flexible tolerance remains unsolved. Additionally, historical map sheets are published every few years. The current methods did not address the problem of predicting the road features in the time between publication years. Also, further and wider applications of machine learning to road feature extraction from historical maps deserve exploration. For example, due to the spatial distortion, the extracted road data need correcting. How to apply machine learning to correct the extracted road data remains unsolved. Further, the corrected road data can be used to correct other map features that have already been extracted, like the alignment of buildings to road lines, as buildings are usually located along roads. The process of alignment could be modeled using a machine learning framework.

### 3.4.2. Current limitations and technology trends

There still exist limitations in the current techniques. Roads are represented by long, slender, and intersected linear features on raster maps. It will benefit road extraction performance by employing high-level semantic information, like multi-scale spatial information. However, current techniques may fail in recognizing and taking advantage of high-level semantic information (e.g., road network structures) (He *et al.*, 2019). Moreover, despite the superiority of machine learning methods, up to now it still remains challenging to automatically generate large amounts of high-quality training data to replace the tedious manual labeling tasks. Usually, machine learning methods are developed only to extract road pixels. Thus, vectorization is performed in a following separate step. It deserves exploration to develop a complete machine learning workflow to directly obtain road vector data from raster maps (e.g., the method that combines the DNN and FSM proposed by Wang *et al.* (2015), the multi-scale machine learning framework developed by Lu *et al.* (2019)). Moreover, the obtained road vector data should be topologically correct to construct road networks. These limitations, however, point to future technology trends, like combining different data sources or using machine learning to automatically generate training data (Duan *et al.*, 2020). In addition, image inpainting involves filling in missing regions of an image (Nazeri *et al.*, 2019). Thus, it can be used to address the quality defects in the historical maps. Furthermore, evaluation metrics that can accurately estimate the topology and continuity of the extracted road features are to be designed. Some methods from the remote sensing domain are hopefully beneficial to address these limitations (e.g., to enable machine learning models to grasp global contextual information by adjusting the receptive field of filters using dilated convolution (Zhou *et al.*, 2018), ASPP (He *et al.*, 2019), or SIIS (Tao *et al.*, 2019); to automatically generate road training data from GPS trajectories (Zhang *et al.*, 2020); to evaluate the connectivity and topology of the extracted roads by comparing the lengths of the corresponding paths in the ground truth and the predictions (Wegner *et al.*, 2015)).

## 3.5. Conclusions

This article provides a detailed review of the studies on road feature extraction from raster maps, which helps to gain a thorough understanding of the existing methods. A novel categorization scheme is proposed to classify the road extraction methods, which associates the image processing techniques with their goals. Moreover, the road extraction methods are characterized by the techniques. Particular attention has been paid to interactive road extraction methods, as they have still been used until recently, but are largely neglected by existing surveys. Furthermore, it has been clarified how the methods in the existing studies have been developed and improved with the continuous advancement of technologies, as well as with the diverting of research hotspots. The issues and directions of further development of road extraction methods are discussed. Machine learning dominates the recent studies. Future research endeavors need to be made to retrieve the accurate and detailed spatial locations of road features.

Moreover, it is necessary to pay special attention to the recent progress in road extraction from overhead imagery, as many such approaches can be tailored and applied to extract features from historical maps.

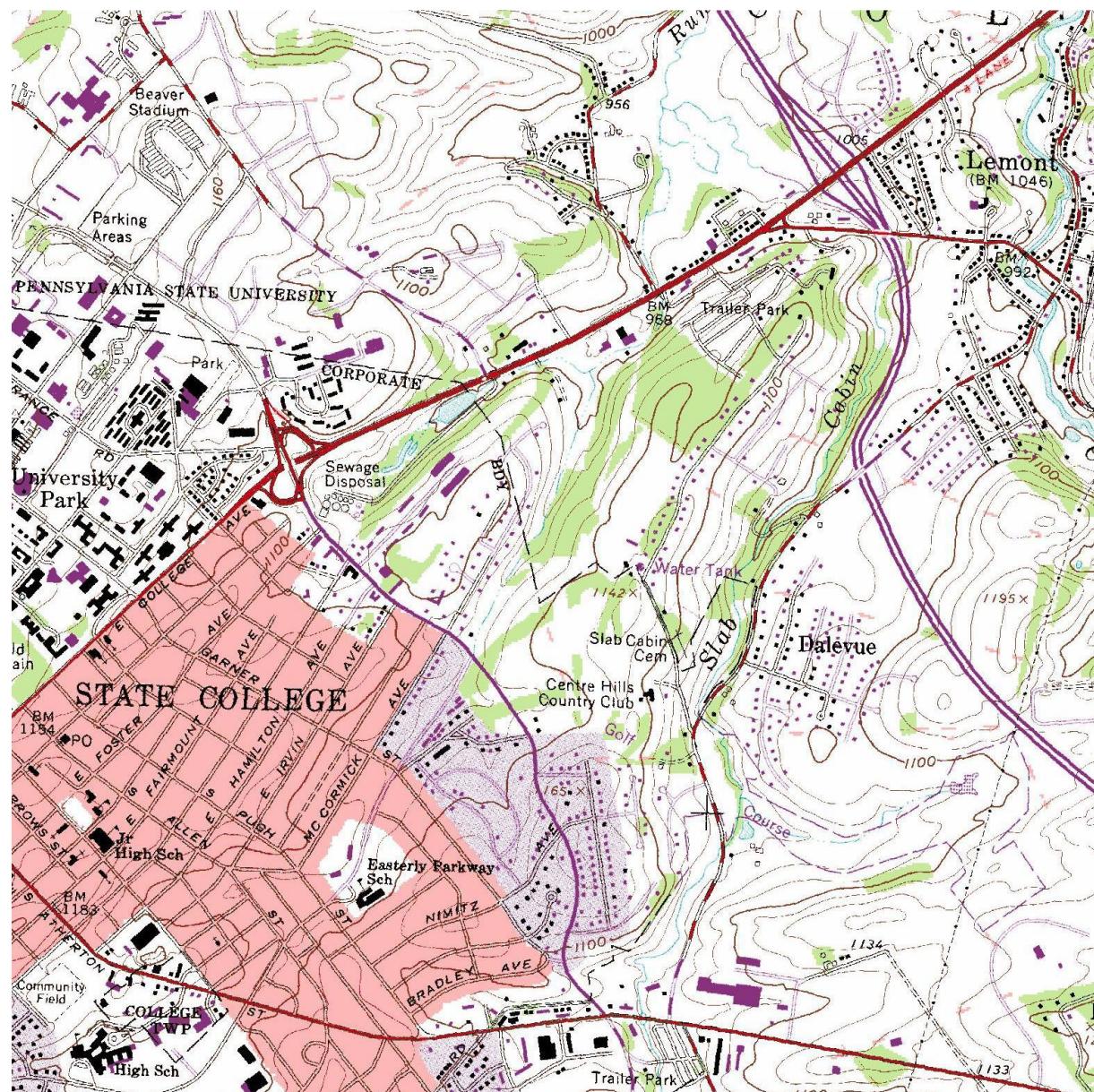
## Appendix

### 1. Word cloud



Figure S3.1. Word cloud of technique terms included in the reviewed papers.

## 2. Figural examples of raster maps



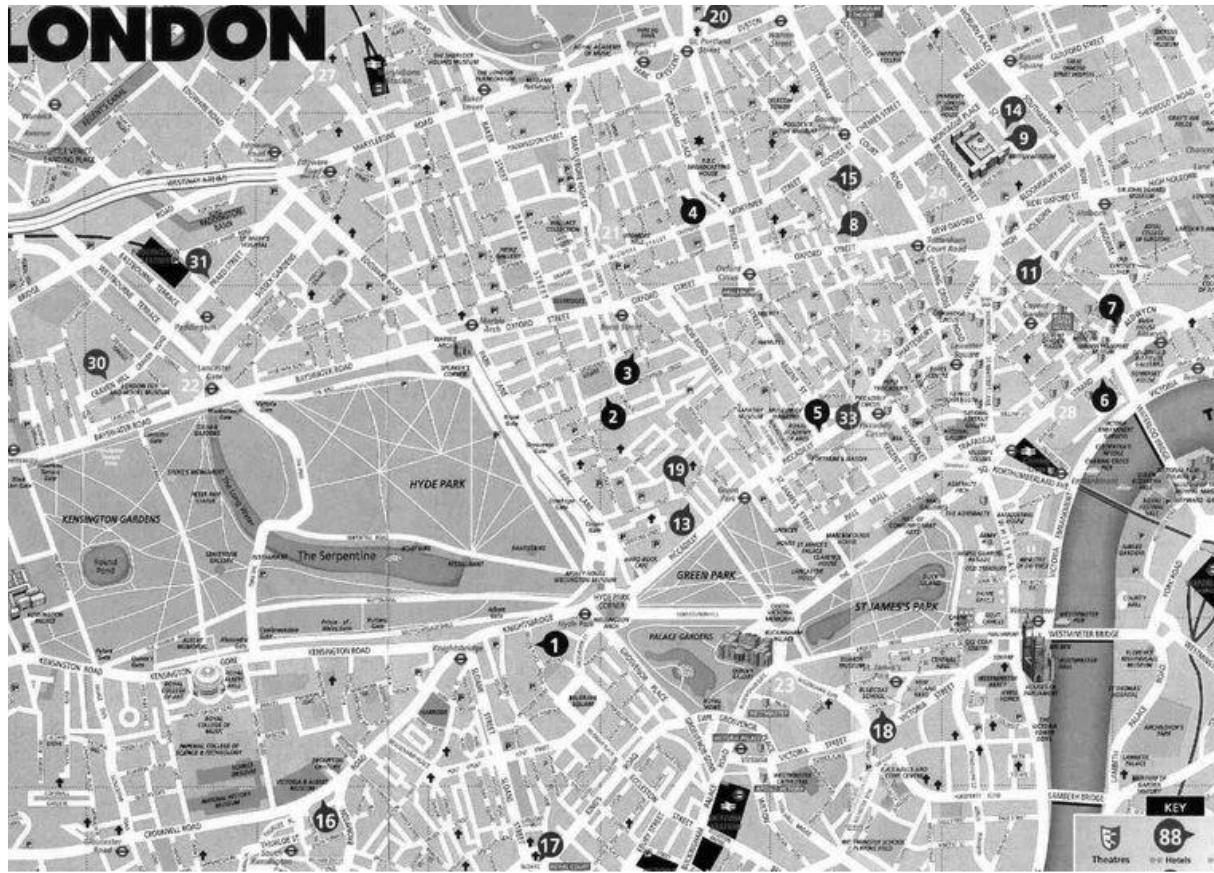


Figure S3.3. A figure from Callier and Saito (2011), showing a scanned map of London.

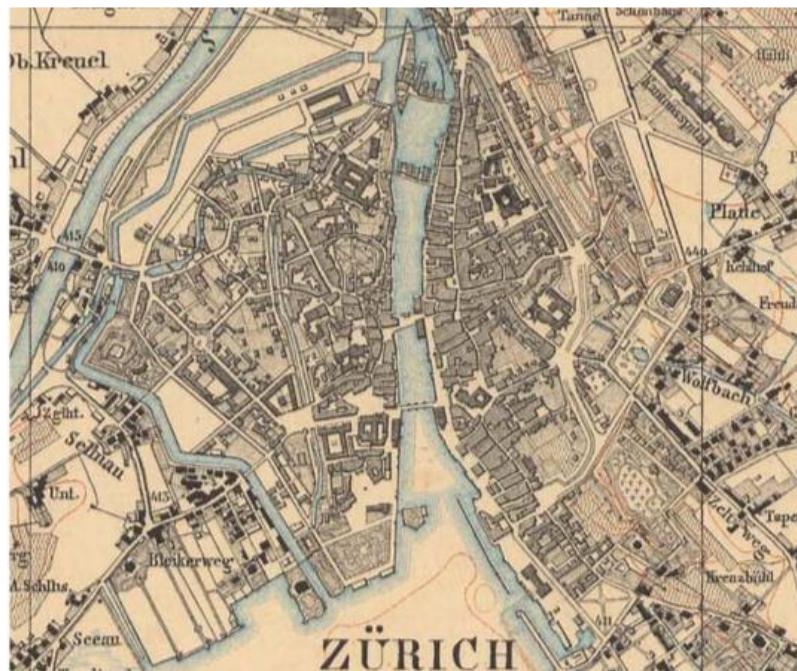


Figure S3.4. A figure from Iosifescu *et al.* (2016), showing a section of the Wild's Topographical Map of Zurich city.

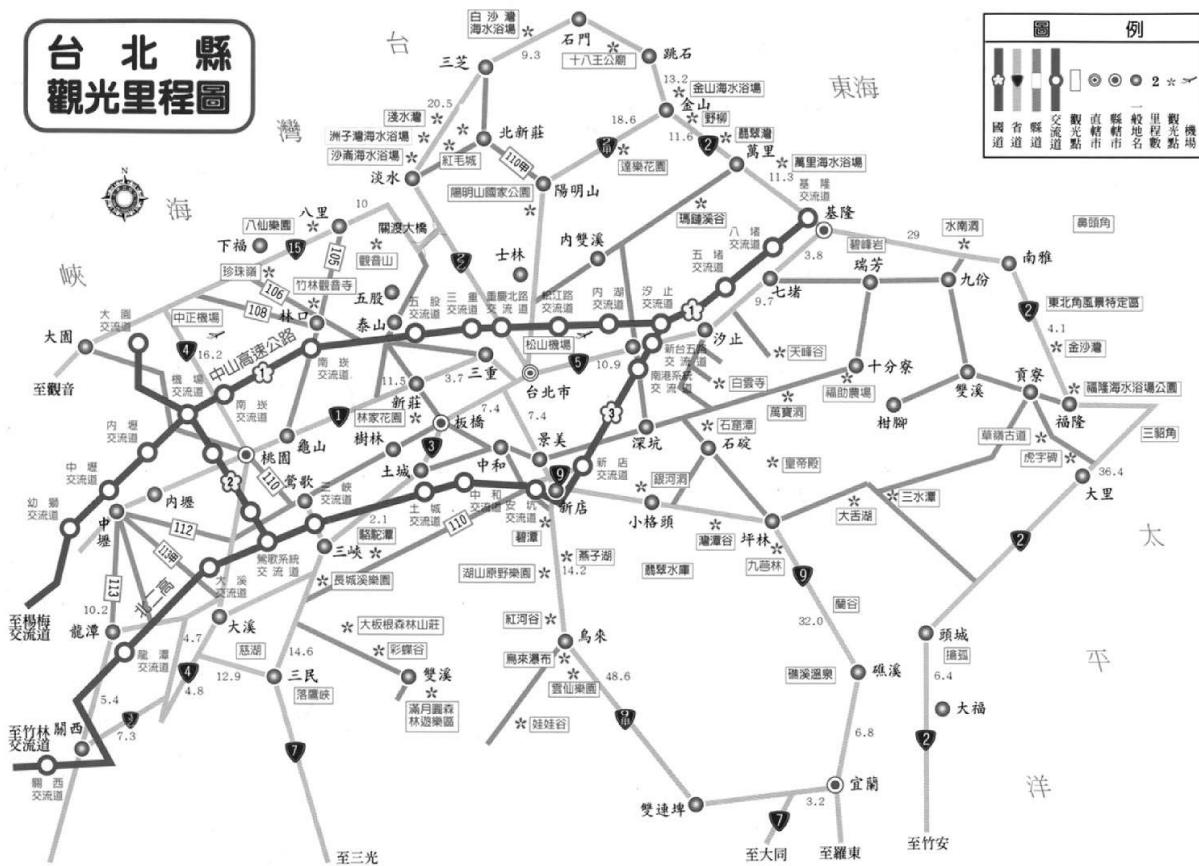
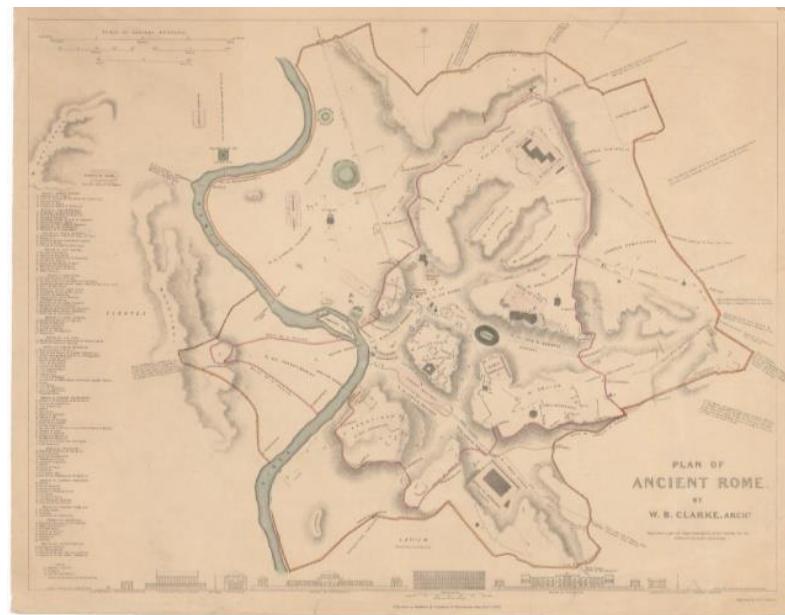
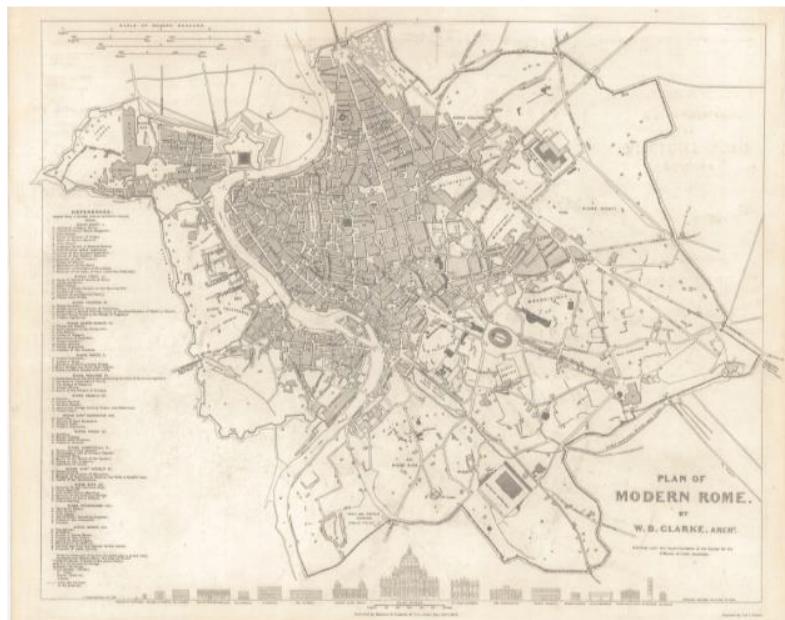


Figure S3.5. A figure from Yin and Huang (2001), showing a scanned Chines road map.



(a)



(b)

Figure S3.6. A figure from Tsorlini *et al.* (2014), showing the Plan of Ancient Rome (a) and Plan of Modern Rome (b).



Figure S3.7. A figure from Dumenie et al. (2019), showing a section of the scanned Cassini map of France (sheet 52, 1759-1777).

## References

- Abdollahi, A., Pradhan, B., Shukla, N., Chakraborty, S. and Alamri, A., 2020. Deep learning approaches applied to remote sensing datasets for road extraction: A state-of-the-art review. *Remote Sensing*, 12(9), p.1444. <https://doi.org/10.3390/rs12091444>
- Ahn, C., Kim, K.S., Rhee, S.B. and Lee, K.Y., 1997, August. A road extraction method from topographical map images. In *1997 IEEE Pacific Rim Conference on Communications, Computers and Signal Processing, PACRIM. 10 Years Networking the Pacific Rim, 1987-1997* (Vol. 2, pp. 839-842). IEEE. <https://doi.org/10.1109/PACRIM.1997.620390>
- Alemany, J. and Kasturi, R., 1988, January. A computer-vision system for interpretation of paper-based maps. In *Applications of Digital Image Processing X* (Vol. 829, pp. 125-139). SPIE. <https://doi.org/10.1117/12.942118>
- Amin, T.J. and Kasturi, R., 1987. Map data processing: recognition of lines and symbols. *Optical Engineering*, 26(4), pp.354-358. <https://doi.org/10.1117/12.7974079>
- Axhausen, K.W., Froelich, P. and Tschopp, M., 2011. Changes in Swiss accessibility since 1850. *Research in Transportation Economics*, 31(1), pp.72-80. <https://doi.org/10.1016/j.retrec.2010.11.010>
- Ballard, D.H., 1981. Generalizing the Hough transform to detect arbitrary shapes. *Pattern recognition*, 13(2), pp.111-122. [https://doi.org/10.1016/0031-3203\(81\)90009-1](https://doi.org/10.1016/0031-3203(81)90009-1)
- Baltsavias, E. and Zhang, C., 2005. Automated updating of road databases from aerial images. *International Journal of Applied Earth Observation and Geoinformation*, 6(3-4), pp.199-213. <https://doi.org/10.1016/j.jag.2004.11.002>
- Ben-Hur, A., Horn, D., Siegelmann, H.T. and Vapnik, V., 2001. Support vector clustering. *Journal of machine learning research*, 2(Dec), pp.125-137. <https://doi.org/10.1162/15324430260185565>
- Bin, D. and Cheong, W.K., 1998, May. A system for automatic extraction of road network from maps. In *Proceedings. IEEE international joint symposia on intelligence and systems (Cat. No. 98EX174)* (pp. 359-366). IEEE. <https://doi.org/10.1109/IJSIS.1998.685476>
- Bovik, A.C. ed., 2009. *The essential guide to image processing*. Academic Press.
- Callier, S. and Saito, H., 2011, June. Automatic Road Extraction from Printed Maps. In *MVA* (pp. 243-246).

Cardim, G.P., Silva, E.A.D., Dias, M.A., Bravo, I. and Gardel, A., 2018. Statistical evaluation and analysis of road extraction methodologies using a unique dataset from remote sensing. *Remote Sensing*, 10(4), p.620. <https://doi.org/10.3390/rs10040620>

Casali, Y. and Heinemann, H.R., 2019. A topological analysis of growth in the Zurich road network. *Computers, Environment and Urban Systems*, 75, pp.244-253.

<https://doi.org/10.1016/j.compenvurbsys.2019.01.010>

Chen, C.C., Knoblock, C.A. and Shahabi, C., 2008. Automatically and accurately conflating raster maps with orthoimagery. *GeoInformatica*, 12, pp.377-410. <https://doi.org/10.1007/s10707-007-0033-0>

Chen, C.C., Knoblock, C.A., Shahabi, C., Chiang, Y.Y. and Thakkar, S., 2004, November. Automatically and accurately conflationg orthoimagery and street maps. In *Proceedings of the 12th annual ACM international workshop on Geographic information systems* (pp. 47-56). <https://doi.org/10.1145/1032222.1032231>

Chen, L.C., Papandreou, G., Schroff, F. and Adam, H., 2017. Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587*.

Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F. and Adam, H., 2018. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European conference on computer vision (ECCV)* (pp. 801-818). [https://doi.org/10.1007/978-3-030-01234-2\\_49](https://doi.org/10.1007/978-3-030-01234-2_49)

Chiang, Y.Y., Duan, W., Leyk, S., Uhl, J.H. and Knoblock, C.A., 2020. *Using Historical Maps in Scientific Studies: Applications, Challenges, and Best Practices*. Cham: Springer. <https://doi.org/10.1007/978-3-319-66908-3>

Chiang, Y.Y. and Knoblock, C.A., 2006, August. Classification of line and character pixels on raster maps using discrete cosine transformation coefficients and support vector machine. In *18th International Conference on Pattern Recognition (ICPR'06)* (Vol. 2, pp. 1034-1037). IEEE. <https://doi.org/10.1109/ICPR.2006.368>

Chiang, Y.Y. and Knoblock, C.A., 2008, November. Automatic extraction of road intersection position, connectivity, and orientations from raster maps. In *Proceedings of the 16th ACM SIGSPATIAL international conference on Advances in geographic information systems* (pp. 1-10). <https://doi.org/10.1145/1463434.1463463>

Chiang, Y.Y. and Knoblock, C.A., 2009a, July. A method for automatically extracting road layers from raster maps. In *2009 10th International Conference on Document Analysis and Recognition* (pp. 838-842). IEEE. <https://doi.org/10.1109/ICDAR.2009.274>

Chiang, Y.Y. and Knoblock, C.A., 2009b. Extracting road vector data from raster maps. In *International Workshop on Graphics Recognition* (pp. 93-105). Springer, Berlin, Heidelberg. <https://doi.org/10.5555/1875532.1875541>

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

Chiang, Y.Y., Knoblock, C.A. and Chen, C.C., 2005, November. Automatic extraction of road intersections from raster maps. In *Proceedings of the 13<sup>th</sup> annual ACM international workshop on Geographic information systems* (pp. 267-276). <https://doi.org/10.1145/1097064.1097102>

Chiang, Y.Y., Knoblock, C.A., Shahabi, C. and Chen, C.C., 2009. Automatic and accurate extraction of road intersections from raster maps. *GeoInformatica*, 13, pp.121-157. <https://doi.org/10.1007/s10707-008-0046-3>

- Chiang, Y.Y., Leyk, S. and Knoblock, C.A., 2011. Efficient and robust graphics recognition from historical maps. In *International Workshop on Graphics Recognition* (pp. 25-35). Springer, Berlin, Heidelberg. [https://doi.org/10.1007/978-3-642-36824-0\\_3](https://doi.org/10.1007/978-3-642-36824-0_3)
- Chiang, Y.Y., Leyk, S. and Knoblock, C.A., 2014. A survey of digital map processing techniques. *ACM Computing Surveys (CSUR)*, 47(1), pp.1-44. <https://doi.org/10.1145/2557423>
- Deng, J., Dong, W., Socher, R., Li, L.J., Li, K. and Fei-Fei, L., 2009, June. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition* (pp. 248-255). Ieee. <https://doi.org/10.1109/CVPR.2009.5206848>
- Dhar, D.B. and Chanda, B., 2006. Extraction and recognition of geographical features from paper maps. *International Journal of Document Analysis and Recognition (IJDAR)*, 8, pp.232-245. <https://doi.org/10.1007/s10032-005-0010-9>
- Duan, W., Chiang, Y.Y., Knoblock, C.A., Jain, V., Feldman, D., Uhl, J.H. and Leyk, S., 2017, November. Automatic alignment of geographic features in contemporary vector data and historical maps. In *Proceedings of the 1st workshop on artificial intelligence and deep learning for geographic knowledge discovery* (pp. 45-54). <https://doi.org/10.1145/3149808.3149816>
- Duan, W., Chiang, Y.Y., Leyk, S., Uhl, J.H. and Knoblock, C.A., 2020. Automatic alignment of contemporary vector data and georeferenced historical maps using reinforcement learning. *International Journal of Geographical Information Science*, 34(4), pp.824-849. <https://doi.org/10.1080/13658816.2019.1698742>
- Dumenieu, B., Chadeyron, J., Cristofoli, P., Perret, J., Jolivet, L. and Baciocchi, S., 2019. Engraved footprints from the past. Retrieving cartographic geohistorical data from the Cassini Carte de France, 1750–1789. *Abstracts of the ICA*, 1, p.68. <https://doi.org/10.5194/ica-abs-1-68-2019>
- Ebi, N., Lauterbach, B. and Anheier, W., 1994. An image analysis system for automatic data acquisition from colored scanned maps. *Machine Vision and Applications*, 7, pp.148-164.
- Eidenbenz, C., Käser, C. and Baltsavias, E.P., 2000. ATOMI-Automated reconstruction of Topographic Objects from aerial images using vectorized Map Information. In *XIX ISPRS Congress, Commission III*. Institute of Geodesy and Photogrammetry, ETH-Hönggerberg. <https://doi.org/10.3929/ethz-a-004334218>
- Eikvil, L., Aas, K. and Koren, H., 1995, August. Tools for interactive map conversion and vectorization. In *Proceedings of 3rd International Conference on Document Analysis and Recognition* (Vol. 2, pp. 927-930). IEEE. <https://doi.org/10.5555/839278.840159>
- Erath, A., Löchl, M. and Axhausen, K.W., 2009. Graph-theoretical analysis of the Swiss road and railway networks over time. *Networks and Spatial Economics*, 9, pp.379-400. <https://doi.org/10.1007/s11067-008-9074-7>
- Fortier, M.F.A., Ziou, D., Armenakis, C. and Wang, S., 2001. Automated correction and updating of road databases from high-resolution imagery. *Canadian Journal of Remote Sensing*, 27(1), pp.76-89. <https://doi.org/10.1080/07038992.2001.10854922>
- He, H., Yang, D., Wang, S., Wang, S. and Li, Y., 2019. Road extraction by using atrous spatial pyramid pooling integrated encoder-decoder network and structural similarity loss. *Remote Sensing*, 11(9), p.1015. <https://doi.org/10.3390/rs11091015>
- He, K., Zhang, X., Ren, S. and Sun, J., 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778). <https://doi.org/10.1109/cvpr.2016.90>

Henderson, T.C. and Linton, T., 2009, July. Raster map image analysis. In *2009 10th international conference on document analysis and recognition* (pp. 376-380). IEEE. <https://doi.org/10.1109/ICDAR.2009.31>

Henderson, T.C., Linton, T., Potupchik, S. and Ostanin, A., 2009, July. Automatic segmentation of semantic classes in raster map images. In *Eighth IAPR International Workshop on Graphics Recognition GREC 2009* (p. 17).

Henderson, T.C., 2014. Road and road intersection extraction. *Analysis of Engineering Drawings and Raster Map Images*, pp.141-179. [https://doi.org/10.1007/978-1-4419-8167-7\\_7](https://doi.org/10.1007/978-1-4419-8167-7_7)

Herold, H., 2015. An evolutionary approach to adaptive image analysis for retrieving and long-term monitoring historical land use from spatiotemporally heterogeneous map sources. *Unpublished PhD dissertation. Dresden, Germany: Technische Universität Dresden*.

Howman, C. and Woodsford, P.A., 1977. *The Laser-scan Fastrak automatic digitising system*. Mapping and Charting Establishment RE.

Iosifescu, I., Tsorlini, A. and Hurni, L., 2016. Towards a comprehensive methodology for automatic vectorization of raster historical maps. *e-Perimetron*, 11(2), pp.57-76.

Itonaga, W., Matsuda, I., Yoneyama, N. and Ito, S., 2003. Automatic extraction of road networks from map images. *Electronics and Communications in Japan (Part II: Electronics)*, 86(4), pp.62-72. <https://doi.org/10.1002/ecjb.10144>

Kakumoto, S., Miyatake, T., Shimada, S. and Ejiri, M., 1983. Development of auto-digitizer using multi-color drawings recognition method. In *IEEE Conference on Computer Vision and Pattern Recognition* (pp. 114-120). Piscataway, NJ: IEEE.

Kasturi, R. and Alemany, J., 1988. Information extraction from images of paper-based maps. *IEEE Transactions on Software Engineering*, 14(5), pp.671-675. <https://doi.org/10.1109/32.6145>

Kasturi, R., Fernandez, R., Amlani, M.L. and Feng, W.C., 1989. Map data processing in geographic information systems. *Computer*, 22(12), pp.10-21. <https://doi.org/10.1109/2.42028>

Kennie, T.J., 2014. *Engineering surveying technology*. CRC Press.  
<https://doi.org/10.1201/9781482269376>

Leyk, S., 2009. Segmentation of colour layers in historical maps based on hierarchical colour sampling. In *International Workshop on Graphics Recognition* (pp. 231-241). Springer, Berlin, Heidelberg. [https://doi.org/10.1007/978-3-642-13728-0\\_21](https://doi.org/10.1007/978-3-642-13728-0_21)

Leyk, S. and Boesch, R., 2010. Colors of the past: color image segmentation in historical topographic maps based on homogeneity. *GeoInformatica*, 14, pp.1-21. <https://doi.org/10.1007/s10707-008-0074-z>

Leyk, S., Boesch, R. and Weibel, R., 2005. A conceptual framework for uncertainty investigation in map-based land cover change modelling. *Transactions in GIS*, 9(3), pp.291-322. <https://doi.org/10.1111/j.1467-9671.2005.00220.x>

Li, J., Hu, Q. and Ai, M., 2018. Unsupervised road extraction via a Gaussian mixture model with object-based features. *International Journal of Remote Sensing*, 39(8), pp.2421-2440. <https://doi.org/10.1080/01431161.2018.1425563>

Linton, T.F., 2009. *Semantic Feature Analysis in Raster Maps* (Doctoral dissertation, School of Computing, University of Utah).

Liu, T., Xu, P. and Zhang, S., 2019. A review of recent advances in scanned topographic map processing. *Neurocomputing*, 328, pp.75-87. <https://doi.org/10.1016/j.neucom.2018.02.102>

- Liu, Y., 2002. An automation system: generation of digital map data from pictorial map resources. *Pattern Recognition*, 35(9), pp.1973-1987. [https://doi.org/10.1016/S0031-3203\(01\)00154-6](https://doi.org/10.1016/S0031-3203(01)00154-6)
- Lu, X., Zhong, Y., Zheng, Z., Liu, Y., Zhao, J., Ma, A. and Yang, J., 2019. Multi-scale and multi-task deep learning framework for automatic road extraction. *IEEE Transactions on Geoscience and Remote Sensing*, 57(11), pp.9362-9377. <https://doi.org/10.1109/TGRS.2019.2926397>
- Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G. and Johnson, B.A., 2019. Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS journal of photogrammetry and remote sensing*, 152, pp.166-177. <https://doi.org/10.1016/j.isprsjprs.2019.04.015>
- Masucci, A.P., Stanilov, K. and Batty, M., 2014. Exploring the evolution of London's street network in the information space: A dual approach. *Physical Review E*, 89(1), p.012805. <https://doi.org/10.1103/PhysRevE.89.012805>
- Mayer, H. and Steger, C., 1998. Scale-space events and their link to abstraction for road extraction. *ISPRS Journal of Photogrammetry and Remote Sensing*, 53(2), pp.62-75. [https://doi.org/10.1016/S0924-2716\(97\)00040-3](https://doi.org/10.1016/S0924-2716(97)00040-3)
- Miyatake, T., 1985. Extraction of roads from topological maps using a parallel extraction algorithm. *Systems and Computers in Japan*, 16(6), pp.77-86. <https://doi.org/10.1002/scj.4690160609>
- Mosinska, A., Marquez-Neila, P., Koziński, M. and Fua, P., 2018. Beyond the pixel-wise loss for topology-aware delineation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3136-3145). <https://doi.org/10.1109/cvpr.2018.00331>
- Moučka, B.M., 2018. Semi-automatic tools for image segmentation. (*Unpublished MS thesis*). Brno, Czech Republic: Masaryk University.
- Nakajima, M., Agui, T. and Iituka, H., 1985. A graphical structure extracting method from an urban map using parallel vector tracers. *Systems and Computers in Japan*, 16(6), pp.39-48. <https://doi.org/10.1002/scj.4690160605>
- Nagao, T., Agui, T. and Nakajima, M., 1988, January. An automatic road vector extraction method from maps. In *9th International Conference on Pattern Recognition* (pp. 585-586). IEEE Computer Society. <https://doi.org/10.1109/ICPR.1988.28299>
- Nagao, T., Agui, T. and Nakajima, M., 1990. Automatic extraction of roads denoted by parallel lines from 1/25,000-scaled maps utilizing skip-scan method. *Systems and computers in Japan*, 21(11), pp.96-105. <https://doi.org/10.1002/scj.4690211110>
- Nazeri, K., Ng, E., Joseph, T., Qureshi, F.Z. and Ebrahimi, M., 2019. Edgeconnect: Generative image inpainting with adversarial edge learning. *arXiv preprint arXiv:1901.00212*.
- Nishijima, M. and Watanabe, T., 1998, January. A cooperative inference mechanism for extracting road information automatically. In *Asian Conference on Computer Vision* (pp. 217-224). Berlin, Heidelberg: Springer Berlin Heidelberg. [https://doi.org/10.1007/3-540-63931-4\\_218](https://doi.org/10.1007/3-540-63931-4_218)
- Pezeshk, A., 2011. *Feature extraction and text recognition from scanned color topographic maps*. (Doctoral dissertation, The Pennsylvania State University).
- Pezeshk, A. and Tutwiler, R.L., 2010, March. Improved multi angled parallelism for separation of text from intersecting linear features in scanned topographic maps. In *2010 IEEE International Conference on Acoustics, Speech and Signal Processing* (pp. 1078-1081). IEEE. <https://doi.org/10.1109/ICASSP.2010.5495342>
- Pezeshk, A. and Tutwiler, R.L., 2011. Automatic feature extraction and text recognition from scanned topographic maps. *IEEE Transactions on Geoscience and Remote Sensing*, 49(12), pp.5047-5063. <https://doi.org/10.1109/TGRS.2011.2157697>

Ronneberger, O., Fischer, P. and Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III* 18 (pp. 234-241). Springer International Publishing. [https://doi.org/10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28)

Saeedimoghaddam, M. and Stepinski, T.F., 2020. Automatic extraction of road intersection points from USGS historical map series using deep convolutional neural networks. *International Journal of Geographical Information Science*, 34(5), pp.947-968.  
<https://doi.org/10.1080/13658816.2019.1696968>

Simonyan, K. and Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Steger, C., 1998. An unbiased detector of curvilinear structures. *IEEE Transactions on pattern analysis and machine intelligence*, 20(2), pp.113-125. <https://doi.org/10.1109/34.659930>

Steger, C., 2000. Subpixel-precise extraction of lines and edges. *International archives of photogrammetry and remote sensing*, 33(3), pp.141-156.

Stevenson, P.J., 1994. *Scanning for automated data conversion of cartographic documents* (Vol. 426). Department of Geodetic Science and Surveying, Ohio State University.

Strano, E., Nicosia, V., Latora, V., Porta, S. and Barthélémy, M., 2012. Elementary processes governing the evolution of road networks. *Scientific reports*, 2(1), p.296. <https://doi.org/10.1038/srep00296>

Sun, T., Di, Z., Che, P., Liu, C. and Wang, Y., 2019. Leveraging crowdsourced GPS data for road extraction from aerial imagery. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 7509-7518). <https://doi.org/10.1109/cvpr.2019.00769>

Suzuki, S., Kosugi, M. and Hoshino, T., 1987. Automatic line drawing recognition of large-scale maps. *Optical Engineering*, 26(7), pp.642-649. <https://doi.org/10.1117/12.7974129>

Suzuki, S. and Yamada, T., 1990. MARIS: Map recognition input system. *Pattern recognition*, 23(8), pp.919-933. [https://doi.org/10.1016/0031-3203\(90\)90137-A](https://doi.org/10.1016/0031-3203(90)90137-A)

Tao, C., Qi, J., Li, Y., Wang, H. and Li, H., 2019. Spatial information inference net: Road extraction using road-specific contextual information. *ISPRS Journal of Photogrammetry and Remote Sensing*, 158, pp.155-166. <https://doi.org/10.1016/j.isprsjprs.2019.10.001>

Tsorlini, A., Iosifescu, I., Iosifescu, C. and Hurni, L., 2014. A methodological framework for analyzing digitally historical maps using data from different sources through an online interactive platform. *e-Perimetron*, 9(4), pp.153-165.

Wang, J., Song, J., Chen, M. and Yang, Z., 2015. Road network extraction: A neural-dynamic framework based on deep learning and a finite state machine. *International Journal of Remote Sensing*, 36(12), pp.3144-3169. <https://doi.org/10.1080/01431161.2015.1054049>

Wang, W., Yang, N., Zhang, Y., Wang, F., Cao, T. and Eklund, P., 2016. A review of road extraction from remote sensing images. *Journal of traffic and transportation engineering (English edition)*, 3(3), pp.271-282. <https://doi.org/10.1016/j.jtte.2016.05.005>

Wang, S., Yu, D., Kwan, M.P., Zhou, H., Li, Y. and Miao, H., 2019. The evolution and growth patterns of the road network in a medium-sized developing city: A historical investigation of Changchun, China, from 1912 to 2017. *Sustainability*, 11(19), p.5307. <https://doi.org/10.3390/su11195307>

Watanabe, T. and Oshitani, T., 2001, September. Parallel recognition of roads from urban maps on generation/verification paradigm of hypotheses. In *Proceedings of Sixth International Conference on Document Analysis and Recognition* (pp. 1225-1234). IEEE.

<https://doi.org/10.1109/ICDAR.2001.953978>

Wei, Y., Wang, Z. and Xu, M., 2017. Road structure refined CNN for road extraction in aerial image. *IEEE Geoscience and Remote Sensing Letters*, 14(5), pp.709-713. <https://doi.org/10.1109/LGRS.2017.2672734>

Wegner, J.D., Montoya-Zegarra, J.A. and Schindler, K., 2013. A higher-order CRF model for road network extraction. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1698-1705). <https://doi.org/10.1109/cvpr.2013.222>

Wegner, J.D., Montoya-Zegarra, J.A. and Schindler, K., 2015. Road networks as collections of minimum cost paths. *ISPRS Journal of Photogrammetry and Remote Sensing*, 108, pp.128-137. <https://doi.org/10.1016/j.isprsjprs.2015.07.002>

Wulamu, A., Shi, Z., Zhang, D. and He, Z., 2019. Multiscale road extraction in remote sensing images. *Computational intelligence and neuroscience*, 2019. <https://doi.org/10.1155/2019/2373798>

Yamada, H., Yamamoto, K. and Hosokawa, K., 1993. Directional mathematical morphology and reformalized Hough transformation for the analysis of topographic maps. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(4), pp.380-387. <https://doi.org/10.1109/34.206957>

Yamada, H., Yamamoto, K. and Nakamura, M., 1990, June. Increasing the performance of MAP (multi-angled parallelism) erosion-dilation for feature extraction through unary/binary operations. In [1990] *Proceedings. 10th International Conference on Pattern Recognition* (Vol. 2, pp. 524-528). IEEE. <https://doi.org/10.1109/ICPR.1990.119419>

Yamada, H., Yamamoto, K., Saito, T. and Matsui, S., 1991. MAP: Multi-Angled Parallelism for feature extraction from topographical maps. *Pattern Recognition*, 24(6), pp.479-488. [https://doi.org/10.1016/0031-3203\(91\)90015-W](https://doi.org/10.1016/0031-3203(91)90015-W)

Yang, Y., An, X. and Huang, L., 2012. Vectorization of linear features in scanned topographic maps using adaptive image segmentation and sequential line tracking. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 39, pp.103-108. <https://doi.org/10.5194/isprsarchives-XXXIX-B4-103-2012>

Yin, P.Y. and Huang, Y.B., 2001. Automating data extraction and identification on Chinese road maps. *Optical Engineering*, 40(5), pp.663-673. <https://doi.org/10.1117/1.1357831>

Zhang, C., 2003. *Updating of cartographic road databases by image analysis* (Doctoral dissertation, ETH Zurich).

Zhang, C., Baltsavias, E. and O'Sullivan, L., 2005. Performance evaluation of ATOMI system for road database updating from aerial film, ADS40, IKONOS and Quickbird orthoimagery. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 36(3/W24), pp.165-170. <https://doi.org/10.3929/ethz-b-000000391>

Zhang, J., Hu, Q., Li, J. and Ai, M., 2020. Learning from GPS trajectories of floating car for CNN-based urban road extraction with high-resolution satellite imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 59(3), pp.1836-1847. <https://doi.org/10.1109/TGRS.2020.3003425>

Zhao, P., Jia, T., Qin, K., Shan, J. and Jiao, C., 2015. Statistical analysis on the evolution of OpenStreetMap road networks in Beijing. *Physica A: Statistical Mechanics and its Applications*, 420, pp.59-72. <https://doi.org/10.1016/j.physa.2014.10.076>

Zhou, L., Zhang, C. and Wu, M., 2018. D-LinkNet: LinkNet with pretrained encoder and dilated convolution for high resolution satellite imagery road extraction. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops* (pp. 182-186). <https://doi.org/10.1109/cvprw.2018.00034>



## 4. A novel data augmentation method to enhance the training dataset for road extraction from Swiss historical maps

---

The following chapter is a reprint of the publication “A novel data augmentation method to enhance the training dataset for road extraction from Swiss historical maps” by Chenjing Jiao, Magnus Heitzler, and Lorenz Hurni, <https://isprs-annals.copernicus.org/articles/V-2-2022/423/2022/isprs-annals-V-2-2022-423-2022.html>, published on 17 May 2022 in *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. Content-wise the chapter is reproduced ‘as is’ in the publication. Only formatting changes and corrections of spelling have been made. Code and data used in this paper can be found through <https://github.com/cjiaoeth/Novel-data-augmentation-method-for-road-extraction-from-Swiss-historical-maps>.

---

### Abstract

Long-term retrospective road data are required for various analyses (e.g., investigation of urban sprawl, analysis of road network evolution). Yet, it is challenging to extract roads from scanned historical maps due to their dissatisfying quality. Although deep learning has been exerting its superiority in image segmentation, its application to road extraction from historical maps is rarely seen in existing studies. Deep learning usually requires quite large amounts of training data, which is time-consuming and tedious to label. Data augmentation can to some extent solve this issue. The existing data augmentation techniques vary each training sample as a whole (e.g., rotation, flipping). But some features or symbols on maps will never occur in practice when they are rotated or flipped (e.g., numbers, labels). To solve this problem and to further improve the diversity of training samples, we propose a novel data augmentation method, which varies the target features instead of the whole training sample. The method is validated by applying it to road extraction from the historical Swiss Siegfried map. The experiment results show the effectiveness of the proposed method.

### 4.1. Introduction

Historical maps contain valuable retrospective spatial information that can be rarely found elsewhere. Many historical map series have been scanned into raster format and made widely accessible (Tsorlini *et al.*, 2014). Long-term road network data are used to analyze the evolution of the road networks (Strano *et al.*, 2012; Zhao *et al.*, 2015) and to realistically reconstruct streetscapes of the past for education, entertainment and research purposes<sup>1</sup>. The wide applications of road data and the image processing challenges due to the poor quality of historical maps (e.g., bleaching, paper distortion, blurring) (Leyk *et al.*, 2005) induce an urgent demand for efficient methods to extract roads from historical maps.

Recently, deep learning has become a research hotspot and has been utilized widely owing to its generalizability. Specifically for image processing tasks, convolutional neural networks (CNN) have become the default choice. Convolutional layers in deep learning architectures take input image (patches) of any size and operate on local input regions based on relative spatial coordinates, unlike fully connected networks, which have fixed dimensions and do not explicitly exploit the spatial characteristics. Thus, Long *et al.* (2015) propose Fully Convolutional Network (FCN) by converting conventional fully connected layers to convolutional layers and supplementing the convolutional layers by successive

deconvolution layers for up-sampling. Apart from this, skip connections are added to combine finer scale predictions and coarser ones. The spatial informative output of FCNs make them a natural choice for end-to-end dense prediction tasks like image segmentation (Buslaev *et al.*, 2018). An improvement to the original FCN has been introduced by Ronneberger *et al.* (2015) in the form of the U-Net architecture. Compared with the original FCN, one important modification in U-Net is that the up-sampling part also has a large number of feature channels corresponding to the down-sampling part, which allows the network to propagate context information to higher resolution layers. Consequently, the up-sampling part is more or less symmetric to the down-sampling part. The down-sampling steps gradually generate increasingly abstract feature maps of the input image, while the up-sampling steps progressively reobtain the dimensions of the input and enable precise localization (Ronneberger *et al.*, 2015). Concatenation operations are used to copy the feature maps of an intermediate step in the down-sampling path to the corresponding step in the up-sampling path, which empowers the network to combine low-level and high-level feature representations.

Saeedimoghaddam and Stepinski (2020) employ deep CNNs for road intersection extraction from USGS historical maps. Although with this method, road intersections represented as both single lines and double lines can be successfully extracted, road branches cannot be extracted, which are essential to the analysis of road network growth and urban sprawl (Masucci *et al.*, 2014). Chiang *et al.* (2020) report a set of experiments for railroad extraction from USGS historical maps to investigate the impact of deep CNN architectures on feature extraction accuracy. Despite of the rapid development and the superiority in image segmentation and feature recognition of deep CNNs, their application to road extraction from historical maps is to some extent limited up to now (Jiao *et al.*, 2021).

Unlike the easy availability of historical maps, it is time-consuming and laborious to manually label the corresponding training data. However, deep learning usually requires large amounts of training data. One solution to this issue is data augmentation, which can be used to enhance the size and quality of training datasets so that performant machine learning models can be trained (Shorten and Khoshgoftaar, 2019). Conventional data augmentation methods are applied on the image patch level, which, for example, flip or rotate the image patch as a whole. This study proposes to use data augmentation on the feature level by rotating or flipping the target features only. It not only avoids the generation of possibly unrealistic training data resulting from rotating or flipping some map features (e.g., labels, numbers, triangulation points), but also improves the diversity of training samples, thereby empowering the deep learning network to learn invariant representations unique to target features (e.g., roads). The effectiveness of the novel data augmentation method is verified by applying it to road extraction from the Swiss Siegfried map.

## 4.2. Data and method

### 4.2.1. Data

The Swiss Siegfried map is a comprehensive Swiss national map series published between 1872 and 1949 at the scales of 1:25,000 (Jura and Swiss plateau) and 1:50,000 (Alps) (Heitzler and Hurni, 2020; Jiao *et al.*, 2020). The map series depicts various geographical features such as buildings, roads, railways, hydrological features, vegetation areas. The Siegfried map sheets are scanned into raster format by Swiss Federal Office of Topography, and georeferenced based on the map frame corner points and the coordinate grid lines (Heitzler *et al.*, 2018). The size of each scanned map sheet is 7,000 pixels  $\times$  4,800 pixels. The resolution of map sheets used in this study is 1.25 m/pixel with a scale of 1:25,000. The map sheet has three colour channels, namely RGB.

Roads are represented by six different symbols, namely single dashed line, single solid line, the combination of a solid line and a parallel dashed line, two parallel lines, a thin line together with a thicker line, and two parallel lines with short strokes in between, as marked by red arrows shown in Figure 4.1. The symbols correspond to different road grades.



Figure 4.1. Roads on Siegfried map. Geodata © Swisstopo

The labelled road data we have at hand only covers Zurich city. The red lines in Figure 4.2 show the labelled data, which are road centerlines. Figure 4.2 (a) is an overview of the data overlaying the corresponding Siegfried map sheets, (b) a part of the data, and (c) buffers of roads in (b), as shown by white areas. The buffers are generated based on road width. For example, the width of roads represented by single solid lines and dashed lines is usually four meters, so the buffer size is two meters.

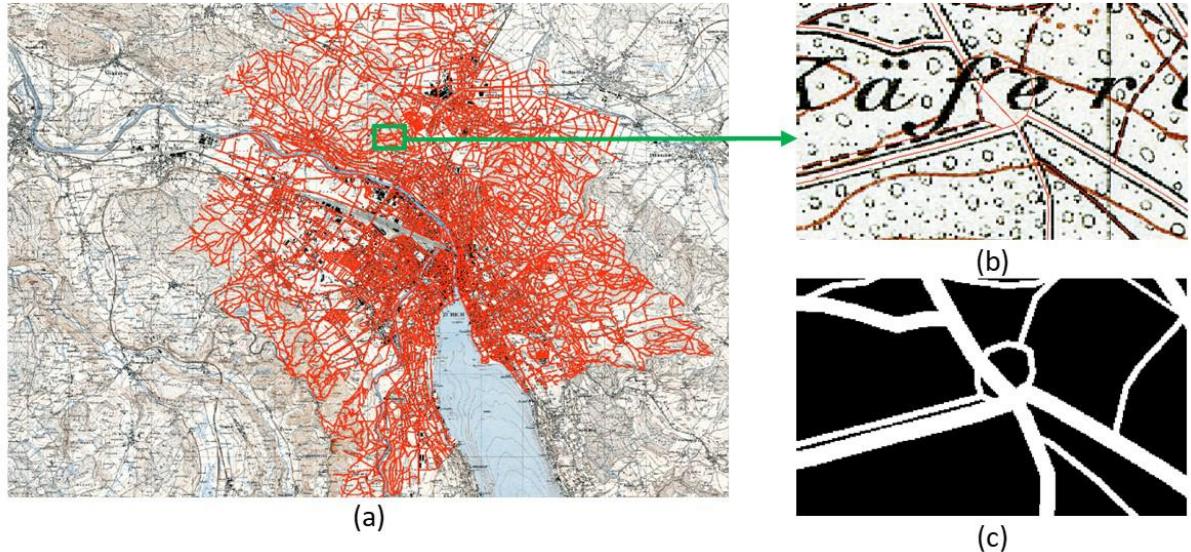


Figure 4.2. The road training data and its buffers. Geodata © Swisstopo

#### 4.2.2. Sampling strategy

To get training samples from the input map sheet and to avoid the data imbalance issue, we adopt the following sampling strategy. First, “positive” points that are located close to roads are randomly generated in road buffers. “Negative” points that are located far from roads are also randomly generated. The positive points are randomly shifted by a small displacement within a neighborhood of 13 pixels  $\times$  13 pixels. Image tiles centered at these sampling points are cropped from the map sheet, which are sized 128 pixels  $\times$  128 pixels. The positive points are shifted as roads will not always go through the center point of an image tile. The green dots in Figure 4.3 show the positive points and red dots the negative ones. The green embossed rectangle represents the map tile cropped centered at one positive point, and the red embossed rectangle the tile centered at one negative point. With this strategy we obtain sampling tiles with roads and without roads, so that the network can learn features of both road areas and non-road areas. In this study, the ratio of the positive samples to the negative samples is empirically set as about 5:1. Additionally, this sampling strategy allows for flexibly adding sampling points for a certain feature (road class in this use case). For example, if we see from the results that a certain road class is not well extracted, sampling points can be added specially for this road class.



Figure 4.3. Sampling points and the corresponding map tiles. Geodata © Swisstopo

#### 4.2.3. A novel data augmentation method

Data augmentation is a data-space solution to the problem of over-fitting as well as limited training data, which are common issues in many applications of deep CNNs (Sun *et al.*, 2019). It encompasses a suite of image transformations, such as scaling, rotation, flipping, colour variation, noise injection, etc. (Shorten and Khoshgoftaar, 2019). Data augmentation enforces the network to learn and identify the desired invariance of feature representations. Specifically for our use case, the learned feature representations of roads should be invariant to variations in the map tiles that are irrelevant for the segmentation task (Dosovitskiy *et al.*, 2014).

Colour and scaling features are essential to road segmentation, so they should not be varied in data augmentation, and there is already much noise in the scanned Siegfried maps. Thus, we use the other two image transformations, namely rotation and flipping. Most of previous data augmentation methods rotate or flip the whole image or image patch. Siegfried maps, however, contain several features that only occur in certain ways. For example, numbers and triangulation points should not be rotated or flipped. Labels can only be rotated slightly and cannot be flipped, as large degree rotation (e.g., larger than 90°) and flipping are not character-preserving transformations (Shorten and Khoshgoftaar, 2019). Therefore, we rotate and flip only the road features, as we have road buffers as ground truth. Specifically, roads are first extracted from the original image patch based on the ground truth. They are randomly rotated or flipped. The remaining features on the patch are not rotated or flipped. The original road areas on the patch are replaced by pixels with the background colour of the Siegfried map. Then, the rotated or flipped roads are overlaid on the patch, which produces an “augmented” patch. The ground truth is also rotated or flipped accordingly. Figure 4.4 shows two examples, where (a) and (c) respectively present the original map tiles cropped from Siegfried map, while (b) displays the result of rotating the roads in (a) by 270° anti-clockwise, and (d) vertical-axis flipping the roads in (c). Especially, the label in (d) is not transformed. These road-only transformations empower the network to learn the features unique to roads, such as long slenderness, colour, connectivity, topology, etc. Furthermore, the road-only rotation and flipping change the relative spatial relation between roads and non-roads, thereby adding more diversity to the training data than previous whole-image (patch) transformations.

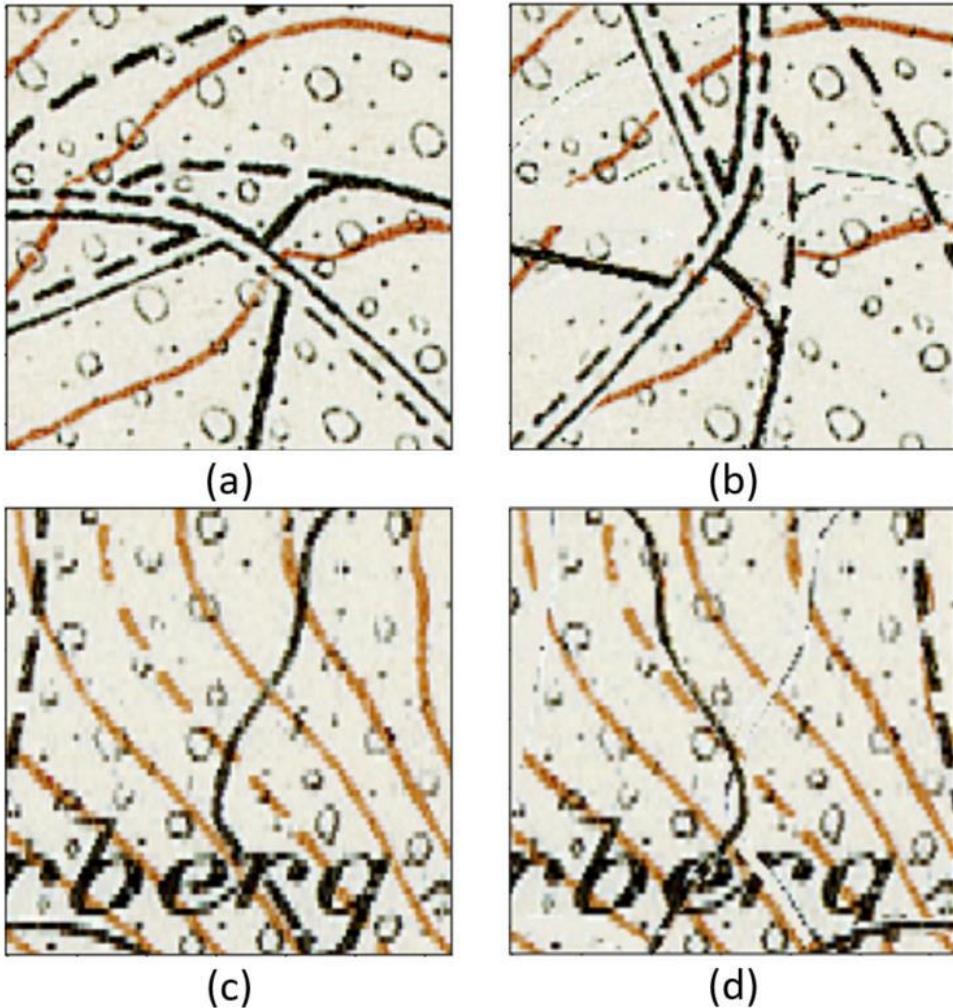


Figure 4.4. (a) a map tile cropped from Siegfried map, (b) rotating the roads in (a) by  $270^\circ$  anti-clockwise, (c) another map tile, (d) horizontal flipping the roads in (c). Geodata © Swisstopo

#### 4.2.4. Road extraction with U-Net

The road segmentation model in this study is developed based on a U-Net architecture. Specifically, the following parameters apply: The first convolution layer of the U-Net in this study has 16 channels. The bottleneck has 256 channels, as each down-sampling step doubles the number of channels. The U-Net consists of four down-sampling steps and four up-sampling steps. Moreover, dropout is used at each down-sampling and up-sampling step with increasing dropout rates towards the bottleneck. It significantly prevents overfitting by avoiding the units co-adapting too much as well as enables to train and combine many different network architectures by randomly sampling a “thinned” network consisting of all the units that survive dropout (Srivastava *et al.*, 2014; Jenny *et al.*, 2020). The network is shown in Figure 4.5. For computing the probability of the produced feature vector being road, a  $1 \times 1$  convolution together with a sigmoid operation (Han and Moraga, 1995) is applied, as shown with a blue rectangle. The target prediction area is sized  $64 \times 64$  pixels, as shown by a yellow embossed square. To enable the model to make precise predictions around the border of the target area, the input map tile is expanded by 32 pixels on each side. Thus, the input tile is  $128 \times 128$  pixels. Furthermore, the weights in the filters are initialized with the method proposed by He *et al.* (2015), which helps with convergence of very deep networks trained directly from scratch. The sampled map tiles and their corresponding road buffers as ground truth are fed into the U-Net for training.

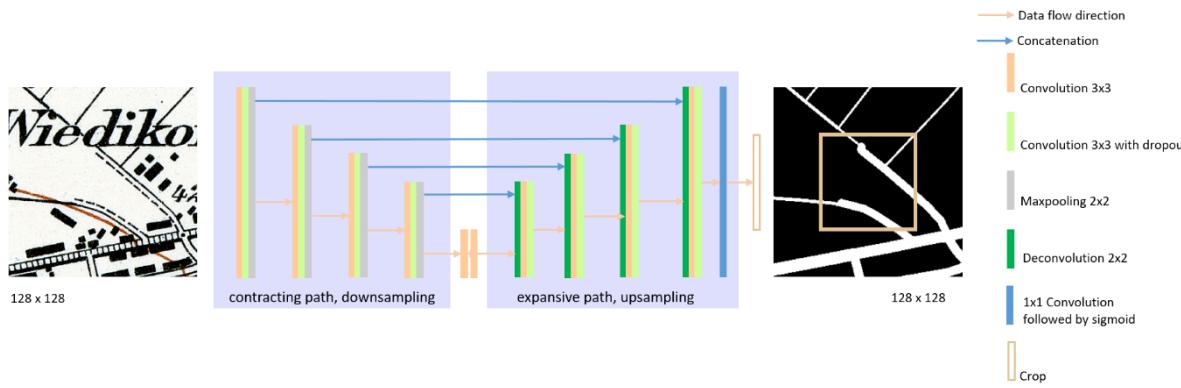


Figure 4.5. The U-Net architecture. Geodata © Swisstopo

## 4.3. Experiment

### 4.3.1. Training scenarios

We use Keras library to implement the experiment. We use Adam optimizer and initialize the learning rate as 0.001 (Kingma and Ba, 2014). Dice loss is used as the loss function (Dice, 1945; Milletari *et al.*, 2016). Each model is trained with 100 epochs. The batch size is 64. To verify the effectiveness of the novel data augmentation method and the flexibility of improving results by adding samples of a certain road class, we implement three training scenarios, namely 1) training with 5000 original samples cropped from the Siegfried map, 2) training with 5000 original samples and additional 1400 samples produced with the novel data augmentation method, and 3) training with 5000 original samples, 1400 samples produced with the novel data augmentation method as well as 500 samples explicitly cropped from features of road class 1. As it is found that road class 1 is less well extracted than other classes, the latter case has been added to specifically improve the extraction capabilities of the model for road class 1, which is represented by dashed line.

### 4.3.2. Post-processing

The trained models are applied to Siegfried map sheets that cover other areas than Zurich city. The pixel values in the raster prediction results indicate the probability of the pixel being a road, as shown with the white areas in Figure 4.6. Pixels with the probability greater than 0.5 are taken as roads. Subsequently, morphological operations are adopted to skeletonize the road areas. Then, the skeletons are vectorized and simplified as road centerlines by the “raster to polyline” tool in ArcGIS. Road extraction results from three typical areas, namely urban area, suburban area and rural area, are reported in Figure 4.7. Red lines represent road centerlines, which overlay the corresponding map images. The overlaid images are shown with 50% transparency to highlight the centerlines.

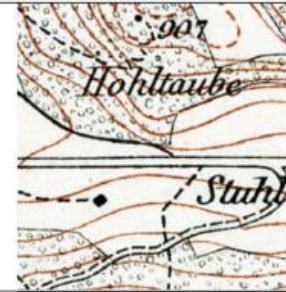
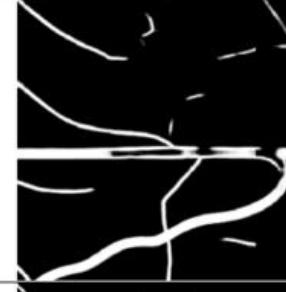
	Town	Rural area
Original map image		
Raster predictions without data augmentation		
Raster predictions with conventional data augmentation		
Raster predictions with novel data augmentation		
Ground truth		

Figure 4.6. The comparison between raster road predictions obtained without data augmentation vs. with conventional data augmentation vs. with novel data augmentation. Geodata © Swisstopo

	Urban area	Suburban area	Rural area
Original map image			
Road centerlines			
Ground truth			

Figure 4.7. Vector road centerlines extracted by the model trained with novel data augmentation.  
Geodata © Swisstopo

#### 4.3.3. Evaluation

As shown in Figure 4.6, raster predictions with conventional and novel data augmentation have much fewer false positives, especially around streams and forest borders, which have very similar shape with roads. Predictions of novel data augmentation are more robust than those of conventional method, especially for double-line roads. In the rural area, the highly curved footpath is extracted with better continuity with the novel method than the conventional method. In addition, we use accuracy and F1 score for quantitatively evaluating raster road predictions, while correctness and completeness for vector road centerlines (Wegner *et al.*, 2013). For raster predictions, the metric values are calculated with the number of correctly or wrongly predicted pixels, namely true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). To compute correctness and completeness, three-meter buffers are generated for the predicted centerlines and the ground truth. Correctness and completeness values are computed with the lengths of TP, TN, FP, FN.

Four Siegfried map sheets, including one covering an area in the northern part of Switzerland, which contains a city and hills (sheet number TA 017 1940), one covering rural area in the Swiss plateau (sheet number TA 199 1941), one covering rural area in the periphery of the Swiss Alps (sheet number TA 385 1941), and the last covering small cities/towns in the eastern part of Switzerland (sheet number TA 219 1944), are selected for evaluating and comparing results obtained by the models trained without data augmentation vs. with novel data augmentation vs. with both novel data augmentation and additional samples around road class 1. The average metrics values of the four sheets are reported in Table 4.1, Table 4.2 and Table 4.3, respectively. Accuracy, F1 and correctness obtained with novel data augmentation outperform those without data augmentation. Especially, correctness is largely improved because of the false positives reduced by applying the novel data augmentation method. Adding additional samples of road class 1 further improves the results, as shown in Table 4.3.

Table 4.1. Quantitative road extraction results (5,000 original samples without data augmentation).

Metrics	Values
Accuracy	96.83%
F1 score	77.37%
Completeness	95.68%
Correctness	64.37%

Table 4.2. Quantitative road extraction results (5,000 original samples plus 1,400 augmented samples of novel data augmentation).

Metrics	Values
Accuracy	97.20%
F1 score	79.63%
Completeness	95.24%
Correctness	66.66%

Table 4.3. Quantitative road extraction results (5000 original samples with additional 500 samples around road class 1, plus 1,400 augmented samples of novel data augmentation).

Metrics	Values
Accuracy	97.76%
F1 score	83.05%
Completeness	96.65%
Correctness	67.35%

#### 4.4. Conclusion

In recent years, deep learning techniques open an avenue to solve the challenge of extracting roads from historical maps. As a solution to the problem of limited training data, data augmentation is commonly used in deep learning applications. To address the fallaciousness due to data augmentation applied to a whole image patch and to enhance the diversity of training samples, this study proposes a novel data augmentation method, which varies the target features instead of the whole image patch. The experiment results show the effectiveness of the proposed method. Especially, the method is very useful to reduce false positives. Although in this study we exemplarily apply the method to road extraction from historical maps, it can be generalizable to other features and data sources. Possible improvements and future work is to explore the optimal ratio of the augmented samples to the original ones.

#### Acknowledgements

We thank Ms. Monika Niederhuber for providing us the road training data of Zurich city as shown in Figure 4.2 (a).

## References

- Buslaev, A., Seferbekov, S., Iglovikov, V. and Shvets, A., 2018. Fully convolutional network for automatic road extraction from satellite imagery. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops* (pp. 207-210). <https://doi.org/10.1109/cvprw.2018.00035>
- Chiang, Y.Y., Duan, W., Leyk, S., Uhl, J.H. and Knoblock, C.A., 2020. *Using historical maps in scientific studies: Applications, challenges, and best practices* (pp. 10-25). Cham: Springer. <https://doi.org/10.1007/978-3-319-66908-3>
- Dice, L.R., 1945. Measures of the amount of ecologic association between species. *Ecology*, 26(3), pp.297-302. <https://doi.org/10.2307/1932409>
- Dosovitskiy, A., Springenberg, J.T., Riedmiller, M. and Brox, T., 2014. Discriminative unsupervised feature learning with convolutional neural networks. *Advances in neural information processing systems*, 27. <https://doi.org/10.1109/TPAMI.2015.2496141>
- Han, J. and Moraga, C., 1995, June. The influence of the sigmoid function parameters on the speed of backpropagation learning. In *International workshop on artificial neural networks* (pp. 195-201). Berlin, Heidelberg: Springer Berlin Heidelberg. [https://doi.org/10.1007/3-540-59497-3\\_175](https://doi.org/10.1007/3-540-59497-3_175)
- He, K., Zhang, X., Ren, S. and Sun, J., 2015. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision* (pp. 1026-1034). <https://doi.org/10.1109/iccv.2015.123>
- Heitzler, M., Gkonos, C., Tsorlini, A. and Hurni, L., 2018. A modular process to improve the georeferencing of the Siegfried map. *e-Perimetron*, 13(2), pp.85-100.
- Heitzler, M. and Hurni, L., 2020. Cartographic reconstruction of building footprints from historical maps: A study on the Swiss Siegfried map. *Transactions in GIS*, 24(2), pp.442-461. <https://doi.org/10.1111/tgis.12610>
- Jenny, B., Heitzler, M., Singh, D., Farmakis-Serebryakova, M., Liu, J.C. and Hurni, L., 2020. Cartographic relief shading with neural networks. *IEEE Transactions on Visualization and Computer Graphics*, 27(2), pp.1225-1235. <https://doi.org/10.1109/tvcg.2020.3030456>
- Jiao, C., Heitzler, M. and Hurni, L., 2020. Extracting Wetlands from Swiss Historical Maps with Convolutional Neural Networks. In *Automatic Vectorisation of Historical Maps. International workshop organized by the ICA Commission on Cartographic Heritage into the Digital 13 March, 2020 Budapest. Proceedings* (pp. 33-38). Department of Cartography and Geoinformatics, ELTE Eötvös Loránd University. <https://doi.org/10.21862/avhm2020.03>
- Jiao, C., Heitzler, M. and Hurni, L., 2021. A survey of road feature extraction methods from raster maps. *Transactions in GIS*, 25(6), pp.2734-2763. <https://doi.org/10.1111/tgis.12812>
- Kingma, D.P. and Ba, J., 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- Leyk, S., Boesch, R. and Weibel, R., 2005. A conceptual framework for uncertainty investigation in map-based land cover change modelling. *Transactions in GIS*, 9(3), pp.291-322. <https://doi.org/10.1111/j.1467-9671.2005.00220.x>
- Long, J., Shelhamer, E. and Darrell, T., 2015. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3431-3440). <https://doi.org/10.1109/cvpr.2015.7298965>

Masucci, A.P., Stanilov, K. and Batty, M., 2014. Exploring the evolution of London's street network in the information space: A dual approach. *Physical Review E*, 89(1), p.012805. <https://doi.org/10.1103/physreve.89.012805>

Milletari, F., Navab, N. and Ahmadi, S.A., 2016, October. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *2016 fourth international conference on 3D vision (3DV)* (pp. 565-571). Ieee. <https://doi.org/10.1109/3dv.2016.79>

Ronneberger, O., Fischer, P. and Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III* 18 (pp. 234-241). Springer International Publishing. [https://doi.org/10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28)

Saeedimoghaddam, M. and Stepinski, T.F., 2020. Automatic extraction of road intersection points from USGS historical map series using deep convolutional neural networks. *International Journal of Geographical Information Science*, 34(5), pp.947-968.

<https://doi.org/10.1080/13658816.2019.1696968>

Shorten, C. and Khoshgoftaar, T.M., 2019. A survey on image data augmentation for deep learning. *Journal of big data*, 6(1), pp.1-48. <https://doi.org/10.1186/s40537-019-0197-0>

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. and Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1), pp.1929-1958.

Strano, E., Nicosia, V., Latora, V., Porta, S. and Barthélémy, M., 2012. Elementary processes governing the evolution of road networks. *Scientific reports*, 2(1), p.296. <https://doi.org/10.1038/srep00296>

Sun, T., Di, Z., Che, P., Liu, C. and Wang, Y., 2019. Leveraging crowdsourced GPS data for road extraction from aerial imagery. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 7509-7518). <https://doi.org/10.1109/cvpr.2019.00769>

Tsorlini, A., Iosifescu, I., Iosifescu, C. and Hurni, L., 2014. A methodological framework for analyzing digitally historical maps using data from different sources through an online interactive platform. *e-Perimetron*, 9(4), pp.153-165.

Wegner, J.D., Montoya-Zegarra, J.A. and Schindler, K., 2013. A higher-order CRF model for road network extraction. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1698-1705). <https://doi.org/10.1109/cvpr.2013.222>

Zhao, P., Jia, T., Qin, K., Shan, J. and Jiao, C., 2015. Statistical analysis on the evolution of OpenStreetMap road networks in Beijing. *Physica A: Statistical Mechanics and its Applications*, 420, pp.59-72. <https://doi.org/10.1016/j.physa.2014.10.076>

## 5. A fast and effective deep learning approach for road extraction from historical maps by automatically generating training data with symbol reconstruction

---

The following chapter is a reprint of the journal paper “A fast and effective deep learning approach for road extraction from historical maps by automatically generating training data with symbol reconstruction” by Chenjing Jiao, Magnus Heitzler, and Lorenz Hurni, <https://www.sciencedirect.com/science/article/pii/S1569843222001716>, published on 22 August 2022 in *International Journal of Applied Earth Observation and Geoinformation*. The content of the chapter is reproduced ‘as is’ in the journal. Only formatting changes and corrections of spelling have been made. Code and data used in this paper can be found through <https://github.com/cjiaoeth/Automatically-generating-training-data-through-symbol-reconstruction>.

---

### Abstract

Historical road data are often needed for different purposes, such as tracking the evolution of road networks, spatial data integration, and urban sprawl investigation. However, road extraction from historical maps is challenging due to their dissatisfying quality, the difficulty in distinguishing road symbols from those of other features (e.g., isolines, streams), etc. Recently, although deep learning, especially deep convolutional neural networks (CNNs), have been successfully applied to extract roads from remote sensing images, road extraction from historical maps with deep learning is rarely seen in existing studies. Apart from this, it is time-consuming and laborious to manually label large amounts of training data. To bridge these gaps, this paper proposes a novel and efficient methodology to automatically generate training data through symbol reconstruction for road extraction. The proposed methodology is validated by implementing and comparing four training scenarios using the Swiss Siegfried map. The experiments show that imitation maps generated by symbol reconstruction are especially useful in two cases. First, if little manually labelled training data are available, models trained on imitation maps alone can already provide satisfactory road extraction results. Second, when training data from imitation maps are mixed with real training data, the resulting models even outperform the models trained on real data alone for some metrics, thus indicating that imitation maps can be a highly valuable addition. This research provides a new insight for fast and effective road extraction from historical maps using deep learning.

### 5.1. Introduction

As a primary data source of historical geographical information, historical maps include abundant features, such as place names, road networks, railways, buildings, vegetation, etc. (Chiang *et al.*, 2020). Nowadays, this historical heritage has been digitalized into raster format through scanning and made widely accessible (Tsorlini *et al.*, 2014). However, historical maps are subject to poor quality, which results from inaccurate surveying and reproduction technologies or chemical and physical deterioration (e.g., bleaching, paper distortion) (Jiao *et al.*, 2021). Moreover, the scanning process could induce blurring and colour aliasing (Leyk *et al.*, 2005; Liu *et al.*, 2019; Uhl and Duan, 2021). Such poor quality issues make historical map processing challenging (Chiang *et al.*, 2014). As an important and typical anthropogenic feature, historical road network plays a pivotal role in many areas, such as road network analysis (Zhao *et al.*, 2015), spatial data integration (Chen *et al.*, 2008), and urban sprawl investigation (Strano *et al.*, 2012), etc. Hence, it is necessary to efficiently extract roads from historical maps.

Before the blossom and wide utilisation of deep learning in image segmentation, various conventional methods are developed for road extraction from historical maps, including line tracing, colour image segmentation (CIS), and morphological operations. For example, in Chiang *et al.* (2005), as well as Chiang and Knoblock (2013), a parallel pattern tracing algorithm is proposed to extract roads from United States Geological Survey (USGS) historical maps. But usually, line-tracing methods may not work for dashed lines. Chiang and Knoblock (2009a) apply CIS techniques for road extraction based on cartographic colour layers. However, CIS may still lack generalisability, as colours vary in different map series or even in the same map series due to quality deficiency of the scanned maps. Moreover, CIS usually demands other techniques as complementary like Hough transform (Chiang and Knoblock, 2013). Morphological operations play an important role in skeletonisation (Chiang and Knoblock, 2009a; Chiang and Knoblock, 2009b), reconnecting broken road lines (Dhar and Chanda, 2006), and refining the extracted road areas or road intersections (Chiang and Knoblock, 2013), etc. A set of directional line detectors adapted from morphological operations are used to extract roads from USGS maps (Pezeshk and Tutwiler, 2011). Nonetheless, some straight segments of large characters on the map were still misclassified as roads. A common issue with the above-mentioned methods is lack of generalisability.

In recent years, deep learning techniques, especially deep CNNs have achieved impressive performance in road extraction from remote sensing images (Chen *et al.*, 2017; Sun *et al.*, 2018; Zhou *et al.*, 2018b; He *et al.*, 2019). However, little attention has been paid to extracting roads from historical maps using deep learning. Compared to roads on remote sensing images presented in their natural form, roads on historical maps are represented by abstract symbols with various shapes and colours (Jiao *et al.*, 2021). Although a tiny number of recent studies attempt to extract road features from historical maps using deep learning, such as road intersection (Saeedimoghaddam and Stepinska, 2020) and railroad (Chiang *et al.* 2020), there is no research to systematically extract road networks from historical maps using deep learning. Moreover, manually labelling training data is usually time-consuming and laborious, whose amount and quality have a great impact on the performance of deep learning models (Jiao *et al.*, 2021). Especially, the complexity of road networks and the demand for enough training data for each grade of roads make the issue even harsher.

To bridge these gaps, we propose a novel methodology to automatically and efficiently generate large amounts of training data, which is used to extract roads from historical maps through deep CNNs. First, considering that the contemporary geo-spatial vector data are easily accessible, we propose to symbolise the contemporary vector data (e.g., roads, railways, buildings) with the symbols of the corresponding features on the historical map to automatically generate training data. Concretely, a map type, we term “imitation map”, is created, which is constituted of contemporary data but is presented with the symbols of the historical map. Second, U-Net (Ronneberger *et al.*, 2015), a typical CNN architecture, is adapted and used to train models for road extraction using the generated training data. Last, four training scenarios are designed to compare the performance of the trained models, namely training with only real maps, training with only imitation maps, and training with mixed real and imitation maps. The whole methodology is validated by applying it to the Swiss Siegfried map to extract roads. The main contributions of this study are as follows. (1) We develop a fast and efficient deep CNN-based approach to systematically extract roads from historical maps. 2 We propose a novel method to automatically generate training data through symbol reconstruction for training deep CNN models. To the best of our knowledge, this is the first study that automatically generates training data by carefully crafting symbols and subsequently applying those to contemporary vector data. (3) We design and implement four training scenarios to validate that our proposed approach can achieve superior prediction performance for certain metrics (e.g. correctness) and if manually produced training data is rare. Note that the easy access of the contemporary vector data makes this methodology generalizable to a wide range of features (e.g., railways, hydro features, vegetation) and to any other raster map series.

## 5.2. Related work

### 5.2.1. Road extraction using deep CNNs

Recently, deep learning techniques, especially CNNs, have been successfully used for road extraction from various data sources. A major part of related studies is focused on remote sensing images. For instance, Sun *et al.* (2018) concatenate two U-Nets to allow for multiple outputs, namely the first U-Net generating structural information, and the second one outputting road segmentation results. Zhou *et al.* (2018b) propose an encoder-decoder CNN, named D-LinkNet, for road extraction from high resolution overhead imagery by inserting dilated convolution layers into the bottleneck to enlarge the receptive field of filters. He *et al.* (2019) integrate atrous spatial pyramid pooling (ASPP) into U-Net to retrieve multi-scale road characteristics. Historical maps serve as another important data source for road extraction (Jiao *et al.*, 2021). For example, Saeedimoghaddam and Stepinski (2020) employ deep CNNs for road intersection extraction from USGS historical maps. Chiang *et al.* (2020) examine different deep CNN architectures for railroad extraction from USGS historical maps.

Despite of blossom of deep CNNs in road extraction, their application to historical maps is still limited (Jiao *et al.*, 2021). Considering that road extraction from overhead imagery is different from that of historical maps, it urgently calls for effective and efficient road extraction methods from historical maps.

### 5.2.2. Automatic generation of training data

One limitation of deep CNNs is the high production cost of training data. One way to alleviate this issue is to fuse existing spatial data to automatically label training data. For instance, Duan *et al.* (2017) align the contemporary spatial vector data to the corresponding features (e.g., roads, water lines) on USGS historical maps to generate training data, and further model the alignment procedure using reinforcement learning (Duan *et al.*, 2020). Similarly, with the contemporary geospatial vector dataset as ancillary data, Uhl *et al.* (2019) automatically sample training data for human settlement extraction from USGS historical maps. Li (2019) builds a conditional generative adversarial network (GAN) to generate synthetic historical map images based on OpenStreetMap (OSM). Then, synthetic text is inserted to the synthetic images as training data for text detection.

Some other methods are developed for automatically generating road training data from overhead imagery, which can possibly be used for raster maps as well (Jiao *et al.*, 2021). In Wang *et al.* (2015), road training data are automatically generated through a vector-guided sampling strategy from very-high-resolution (VHR) overhead imagery. Zhang *et al.* (2020) use GPS trajectory data of cars as training data to extract roads from high resolution overhead imagery. Nevertheless, the small paths that are not frequently visited by cars in the remote rural areas could not be covered in the training data.

In summary, the existing methods to automatically generate training data are either too complicated (e.g., to model the process with reinforcement learning or GAN) or could not work for all the areas (e.g., rural areas).

### 5.3. Data

#### 5.3.1. Siegfried map and the symbols in the black layer

The Swiss Siegfried map is a comprehensive Swiss topographical map series published between 1872 and 1949 (Heitzler and Hurni, 2020; Jiao *et al.*, 2020). The map series depicts geographical features including buildings, roads, railways, hydrological features, contour lines, toponyms, etc. The Siegfried map sheets have been scanned into raster format by Swiss Federal Office of Topography (Swisstopo1), and georeferenced (Heitzler *et al.*, 2018). Each scanned map sheet is sized  $7,000 \times 4,800$  pixels. The map sheets used in this paper are published for scale 1:25,000 and scanned with a spatial resolution of 1.25 m/pixel.

Generally, roads in the map are represented by six different symbols depending on their grades, as shown in Figure 5.1. The single dashed line, single solid line, and a dashed line together with a solid line represent three types of paths, as shown in Figure 5.1(a)-(c). Two parallel lines represent roads that are 3-5 m wide, a thin line in combination with a thick line represents roads wider than five meters, and the composite symbol of a thin line, a thick line, as well as short strokes in between denotes tramways, as shown in Figure 5.1(d)-(f) respectively.

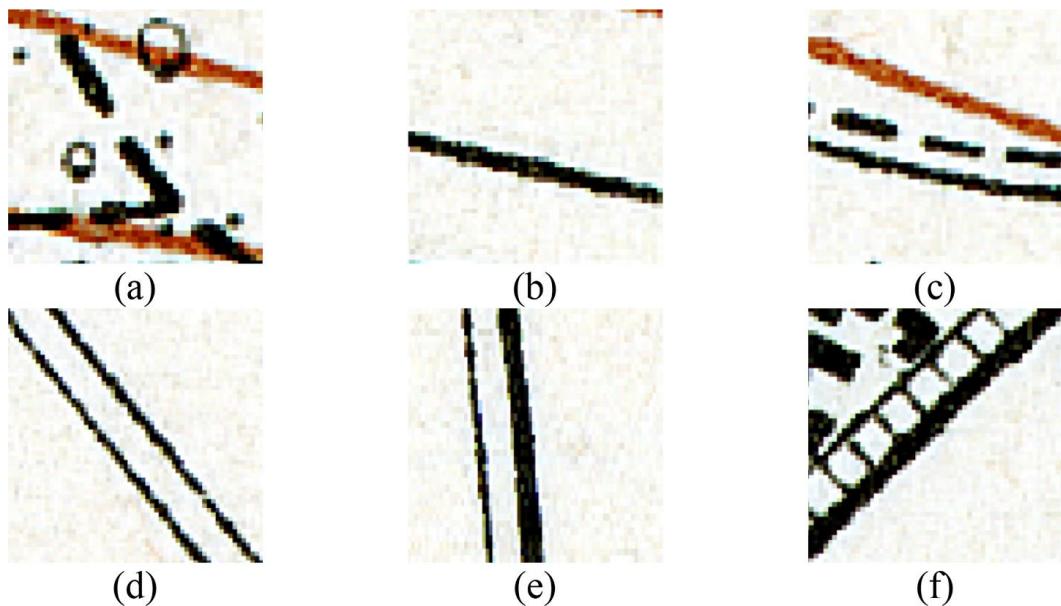


Figure 5.1. Different grades of roads are represented by six types of black line symbols: (a) footpath, (b) field path, (c) unpaved road (d) a 3–5-meter-wide road, (e) a road wider than five meters, (f) a tramway. Geodata © Swisstopo

Other geographical features represented as black symbols are shown in Figure 5.2. The black polygons and the grid-distributed small dots respectively represent buildings and gardens in Figure 5.2(a). The thick black parallel lines in Figure 5.2(b) represent railways. Labels, elevation number, triangulation points symbolized as a triangle with a dot, and coordinate grid lines represented by a horizontal line and a vertical line are shown in Figure 5.2(c). Forests are symbolized as a set of clustered black circles and black dots in Figure 5.2(d). District boundaries, municipality boundaries and cantonal boundaries are symbolized as linear shaped composite points or line segments, as shown in Figure 5.2(e)-(g). Vineyards

are symbolized as vertical strokes distributed in array in Figure 5.2(h). Apart from these features, the map sheet frame is represented as horizontal and vertical black lines.

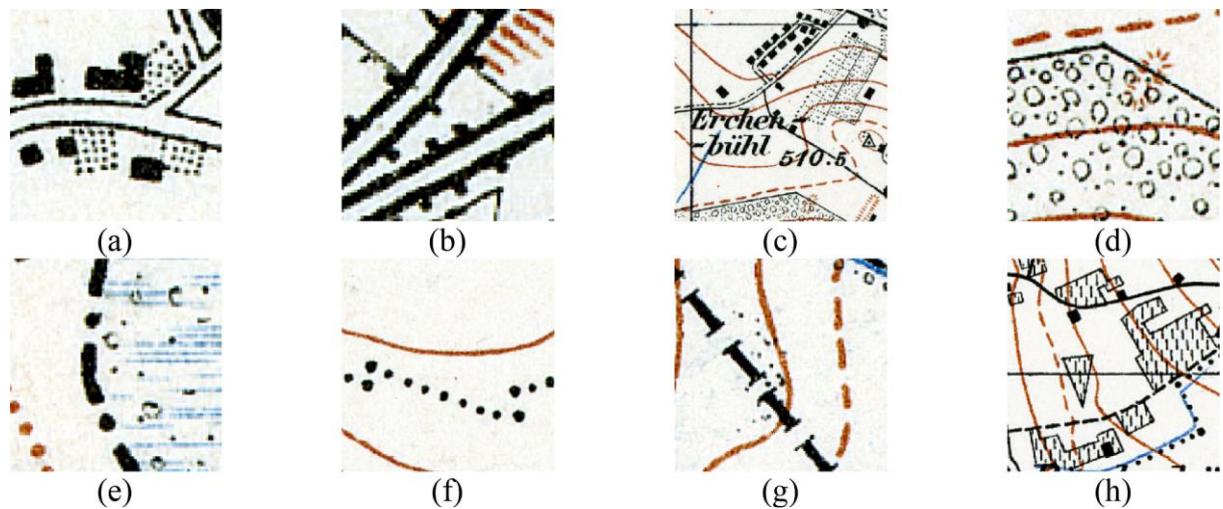


Figure 5.2. Other features represented by black symbols, (a) buildings and gardens, (b) railways, (c) labels, elevation number, triangulation point and coordinate grid lines, (d) forest and its boundary, (e) district boundary, (f) municipality boundary, (g) cantonal boundary, (h) vineyard. Geodata © Swisstopo

### 5.3.2. The available road training data

Only a limited amount of historical road vector data are available for training, which cover Zurich city, as shown in Figure 5.3. Zurich city is covered by the four Siegfried map sheets. The data are road centerlines of the year 1940, which are presented as red lines and overlaying the corresponding Siegfried maps. The data include the six road grades illustrated in Section 5.3.1. A small section of the data in the suburban area and another small section in the urban area are shown in Figure 5.3(b) and (c), respectively.

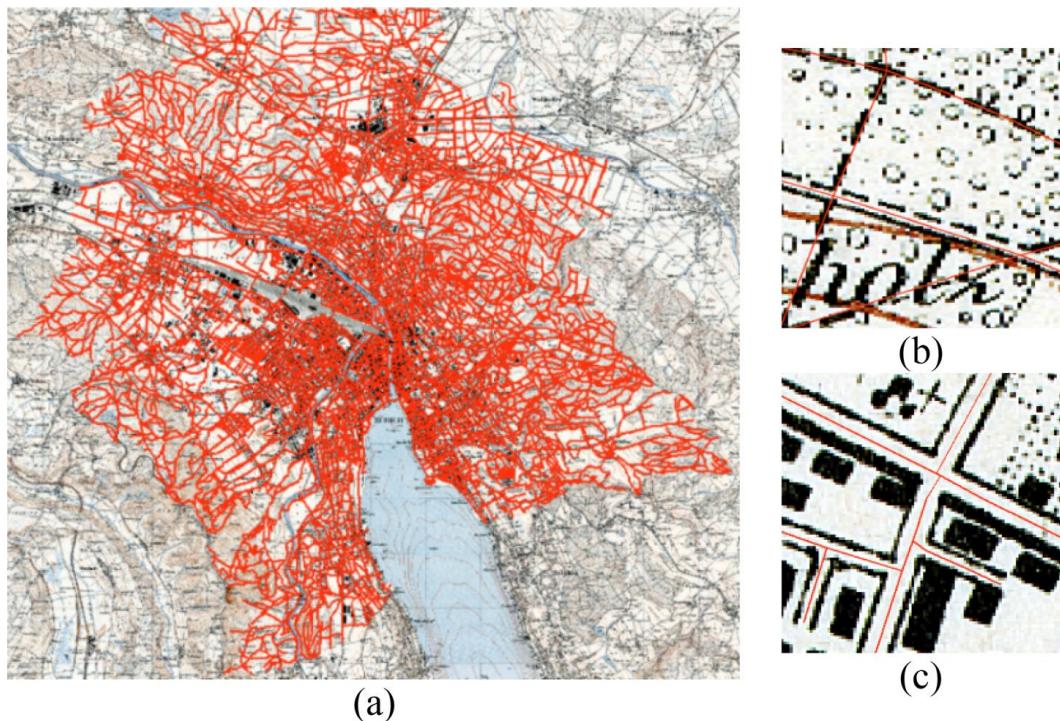


Figure 5.3. The historical road vector data available for training, (a) the whole dataset, (b) roads in the suburban area, (c) roads in the urban area. Geodata © Swisstopo

### 5.3.3. The contemporary vector data

The contemporary vector data that are used to create the imitation Siegfried map are obtained from the swissTLM3D<sup>38</sup> dataset. SwissTLM3D is “the large-scale topographic landscape model of Switzerland”. It contains spatial information (e.g., position, shape) and non-spatial information (e.g., the landscape feature type, construction year of artificial features, place name). The dataset comprises eight topics, namely road networks, public transport, buildings, land usage, land cover, hydrography, single point objects (e.g., the triangulation points) and names (e.g., city names, mountain names).

We use the 2020 swissTLM3D data to create the imitation Siegfried map by symbolizing the data in feature classes corresponding to those in the Siegfried map. For roads, five out of the six road grades stated in Section 5.3.1 (except for tramways) are derived according to the attribute “object type” from the feature class “road”. Tramway and railway are retrieved according to the attributes “object type” and “on street” from the feature class “railroad”. More details on associating the “object type” value with corresponding road symbol are displayed in the Appendix Table S5.1. Building footprints are obtained from the “building” feature class. Gardens and vineyards are retrieved from the “land use” feature class, and forests from “land cover” feature class. Labels and administrative borders are obtained from the “place name” feature class, height from the “height” attribute, and triangulation points from the feature class “single object”. The map frame is retrieved from the 1:25,000 overview data, and coordinate gridlines are placed as a kilometer grid.

---

<sup>38</sup> <https://www.swisstopo.admin.ch/en/geodata/landscape/tlm3d.html> (last access on 29.Oct.2023)

## 5.4. Methodology

In this section, we elaborate on the proposed methodology, as shown in Figure 5.4. First, the symbols in (b) are generated from the Siegfried map (a) via symbol reconstruction. Second, the imitation map in (d) is created by symbolizing the contemporary vector data in (c) with the corresponding reconstructed symbols. Third, the imitation maps and the ground truth (i.e., the road buffers in this case, which are buffered around road lines in the swissTLM3D dataset according to road widths) in (e) from the contemporary dataset are fed into the CNN model for training. Then, the trained models (f) are applied to the black layer of the Siegfried map in (g), which is separated from the original Siegfried map through CIS techniques. Last, the road prediction results in (h) are obtained. Key steps are depicted in the following sections.

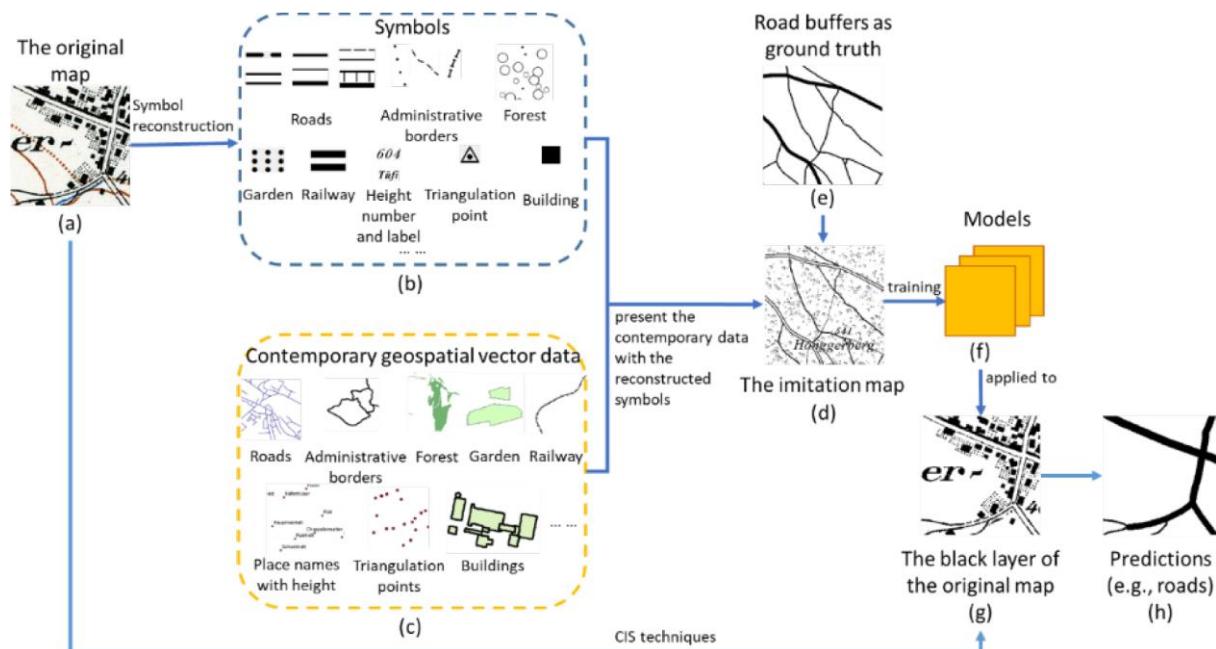


Figure 5.4. Workflow of the methodology. Geodata © Swisstopo

### 5.4.1. Symbol reconstruction to automatically generate training data

To generate the imitation map using the contemporary dataset, the first step is to reconstruct the symbols on the Siegfried map. Especially, as all roads on the Siegfried map are black, we narrow down the target features for symbolization to the black features. The road symbols shown in Figure 5.1 are reconstructed with the “symbol editor” tool in ArcGIS by editing the provided basic symbols. Road lines are classified and symbolized according to the attribute value of “object type” and the metadata<sup>39</sup> of the swissTLM3D dataset, which correspond to the six categories in Figure 5.5. The symbols for railways, administrative boundaries, forests, gardens, vineyards, triangulation points, and coordinate gridlines are created in the same way. Labels are generated by representing the place names with the same font and size as those in Siegfried map, and some of them are put inclined. In a similar way, the elevation numbers derived from

<sup>39</sup> [https://www.swisstopo.admin.ch/content/swisstopo-internet/de/geodata/landscape/tlm3d/\\_jcr\\_content/contentPar/tabs\\_copy/items/dokumente/tabPar/downloadlist/downloadItems/441\\_1618774195491.download/2021\\_04\\_swissTLM3D\\_1\\_9\\_OK\\_d\\_bf.pdf](https://www.swisstopo.admin.ch/content/swisstopo-internet/de/geodata/landscape/tlm3d/_jcr_content/contentPar/tabs_copy/items/dokumente/tabPar/downloadlist/downloadItems/441_1618774195491.download/2021_04_swissTLM3D_1_9_OK_d_bf.pdf). (last access in 2022)

the “height” attribute of the “place name” feature class are presented, as shown in Figure 5.6. Buildings are symbolized as black areas. Although we use ArcGIS to create the symbols, any other GIS, cartography or graphics software that is able to construct complex symbols can be used for this purpose. By representing the contemporary vector data with the corresponding symbols and rasterizing the representation, the imitation map is created in the form of black-white image, which has the same spatial reference as the Siegfried map.

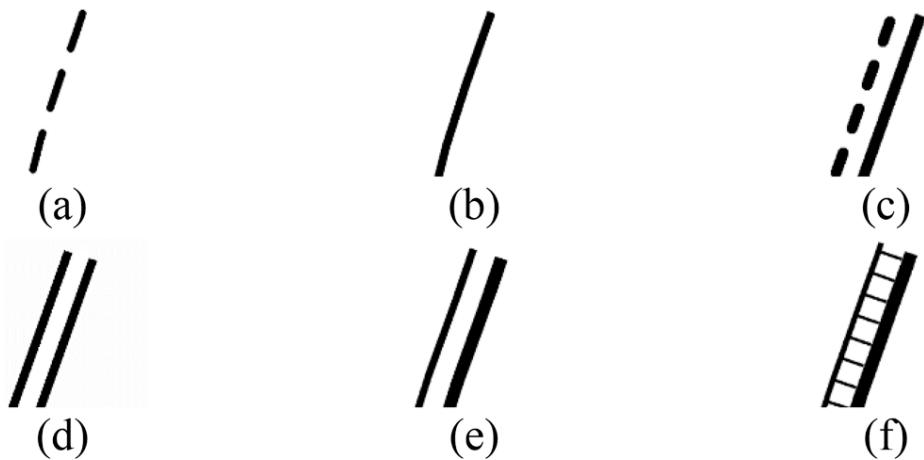


Figure 5.5. Reconstructed road symbols, which correspond to those in Figure 5.1.



Figure 5.6. The spot height and labels derived from the “place name” feature class, which are represented with the same font and size as those of Siegfried map.

Siegfried map and the black layer of the map contain random noise due to blurring and colour aliasing, as shown in Figure 5.7. Thus, to make the imitation map resemble the Siegfried map more and make the trained models more robust (Yousefi-Azar and Hamey, 2017), we add random noise to the grid lines, road lines, railways, buildings, which are represented as white pixels. The noise is added by first generating random points, and then symbolizing each point as a white square sized  $1.25 \text{ m} \times 1.25 \text{ m}$ . For areal features (e.g., buildings), the density of the points is proportional to the area of each feature.

The proportion is determined empirically. We did not set a minimal distance between the points, thereby preserving the randomness of the point distribution (e.g., two points adhere to each other actually present a “double-sized” irregular-shaped noise). To make the imitation map comparable to the Siegfried map, its resolution is set as 1.25 m/pixel. This is why the square of the noise is sized 1.25 m × 1.25 m, as in this way, the random noise is rasterised as a white pixel in the imitation map. Each imitation map sheet is clipped to the same extent as the corresponding Siegfried map sheet, sized 7,000 × 4,800 pixels. The imitation map is further rendered as an RGB image, namely the black pixel value as (0, 0, 0), and the white one as (255, 255, 255). Figure 5.8(a) shows an example of the final imitation map in Zurich. Figure 5.8(b) and 5.8(c) display a part of the urban and rural areas, respectively. Note that the imitation map can be used alone or in combination with the real Siegfried map for training.

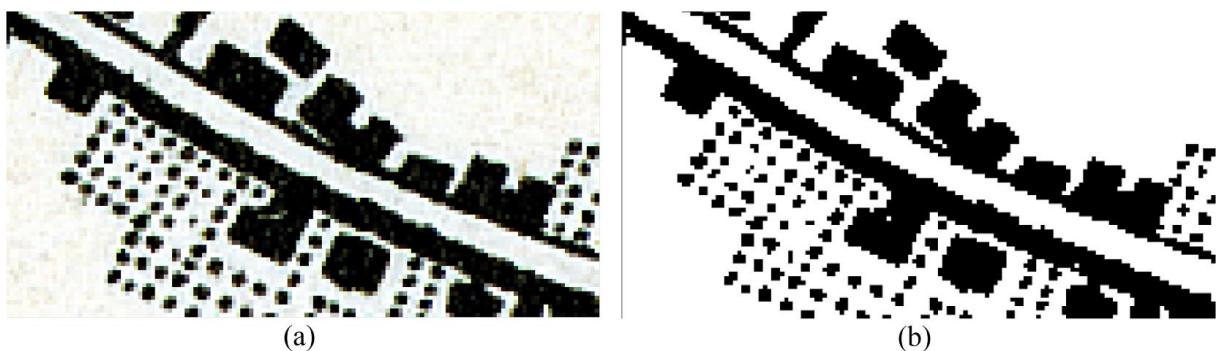


Figure 5.7. (a) Random noise in buildings and road edges on Siegfried map, (b) random noise (white pixels) in the black layer (the method to obtain the black layer is elaborated in Section 5.4.3). Geodata © Swisstopo



Figure 5.8. An example of the imitation map in Zurich, where (a) is an overview, (b) shows the label, tramways, roads, buildings, and a segment of a railway in the urban area, (c) presents a segment of the municipality boundary, paths, a triangulation point, forest, a vineyard, and grid lines in the rural area.

### 5.4.2. Road extraction with U-net

#### 5.4.2.1. Network architecture

The road segmentation model in this paper is developed based on U-Net, as shown in Figure 5.9. For clarity and concision, the repeated operations in the contracting path are referred to as encoder blocks and those in the expansive path as decoder blocks. The encoder block comprises a  $3 \times 3$  convolutional layer followed by an elu (Exponential Linear Unit) activation function (Clevert *et al.*, 2015), a second convolutional layer with dropout (Srivastava *et al.*, 2014) and a  $2 \times 2$  max pooling layer. Elu keeps the identity for positive inputs and produces non-zero values for negative ones. The non-zero values are computed using an  $\alpha$  parameter, and by default  $\alpha = 1.0$ . It allows for faster and more precise learning (Clevert *et al.*, 2015). The decoder block comprises a  $2 \times 2$  transposed convolution to up-sample the feature maps, a concatenation with the corresponding feature maps from the contracting path, a  $3 \times 3$  convolution followed by an elu, a second  $3 \times 3$  convolution with dropout followed by an elu (Ronneberger *et al.*, 2015). In this case, the number of channels in the first and second convolution layer is 16. At each down-sampling step, the number of channels is doubled. At the bottleneck, the number of channels is 256. The up-sampling operation in the decoder blocks doubles the size of the feature maps and “halves the number of channels” (Ronneberger *et al.*, 2015). A  $1 \times 1$  convolution together with a sigmoid operation (Han and Moraga, 1995) is used as a final step to compute the probability of each 16-component feature vector in the produced feature maps being road. To enable the network to “look beyond” the target prediction area, which is shown with a sand-yellow hollow square overlaying the binary ground truth image, the corresponding input map tile is enlarged with a 32-pixel-padding on each side. In this case, the input map tile is  $128 \times 128$  pixels, while the corresponding target prediction area is  $64 \times 64$  pixels.

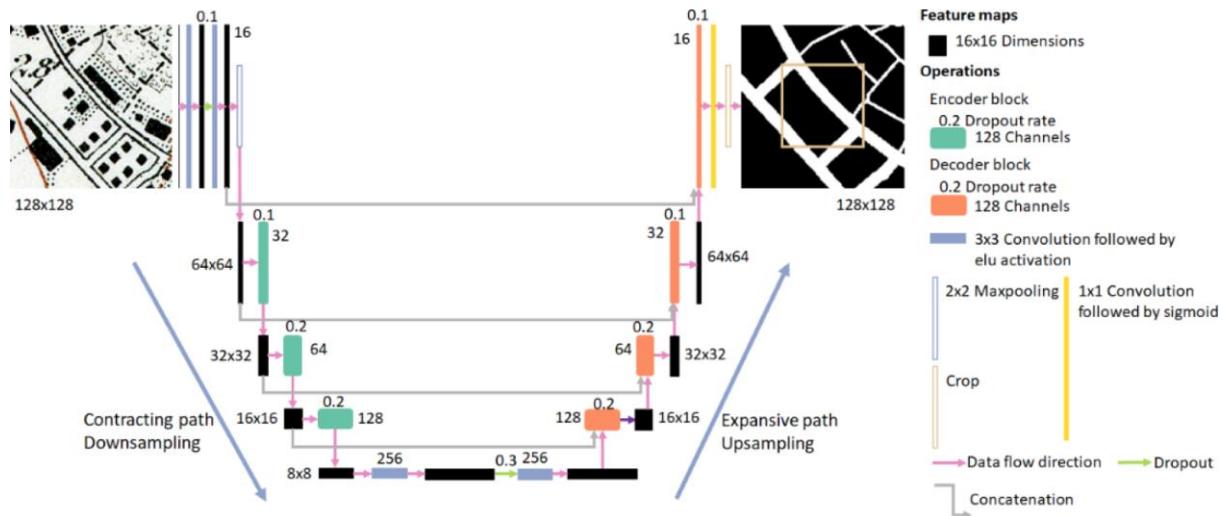


Figure 5.9. The road segmentation model adapted from U-Net. Geodata © Swisstopo

### 5.4.2.2. Training

#### 5.4.2.2.1. Ground-truth preparation

To prepare the ground truth (like the binary image shown in Figure 5.9), the historical road vector data presented in Section 5.3.2 are buffered with a width corresponding to the road grade. As for the contemporary road data, the buffer size (Figure 5.4(e)) is derived from the attribute “object type” and the road grade they belong to. More details on determining buffer sizes of the historical road data and the contemporary road data are presented in Table S5.2 and Table S5.3 of the Appendix. Both road buffers are then rasterized into a binary image with 1.25 m/pixel resolution as the Siegfried map.

#### 5.4.2.2.2. Sampling strategy

To avoid the data imbalance issue, we adopt the following sampling strategy. First, sampling points are randomly generated along road lines. Each point is shifted randomly by a small displacement in the neighborhood of  $13 \times 13$  pixels to avoid roads being always at the center of a patch. Moreover, to enable the network to learn what non-road areas look like, “negative” sampling points are generated in the areas distant from road lines or with sparse road lines. Second, a  $128 \times 128$  sub-image centered at each sampling point is cropped from the input map as a training sample. In this way, both sample patches with and without roads are obtained and fed into the segmentation model for training. Figure 5.10 presents the sampling strategy, where the green dots in (a) are the positive points that locate very close to roads, (b) is an image patch centered at one of the points cropped from the map, the red dots in (c) are negative points and (d) is an image patch centered at one of the points. 80% of the samples are used for training, 10% for validation, and 10% for testing.

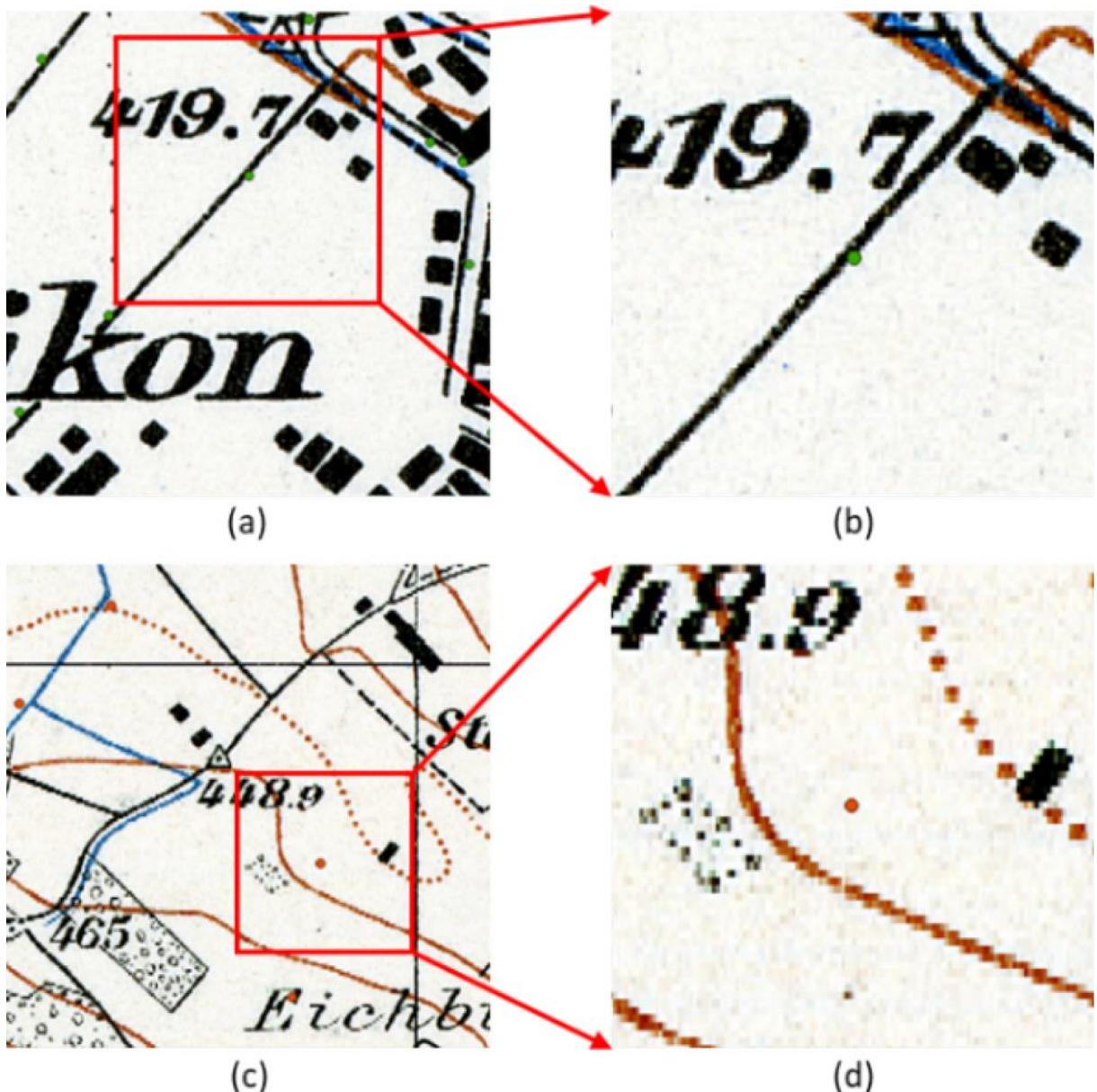


Figure 5.10. The sampling strategy. (a) Positive points, (b) an image patch centered at one positive point, (c) negative points, (d) an image patch centered at one negative point. Geodata © Swisstopo

#### 5.4.2.2.3. Four training scenarios

To verify the effectiveness of the imitation maps as training data, we design four training scenarios to compare the performance of real maps and imitation maps, namely training with only the real maps, training with only the imitation maps, and training with mixed real and imitation ones. The aforementioned four map sheets that cover Zurich city are used for training (see Section 5.3.2). The ground truth road data for the four scenarios are obtained using the method described in Section 5.4.2.2.1. The ratio of the positive samples to the negative samples is about 5:1, which is identical in the four scenarios. Furthermore, to investigate the impact of the number of samples on the training performance, we vary the number of training samples, namely 20,000, 10,000 and 5,000, and apply them to all the four training scenarios. The training scenarios are introduced as follows.

(1) Training with real Siegfried map only.

a. Training on the original Siegfried map (Scenario 1a).

b. Training on the black layer of the Siegfried map (Scenario 1b).

As the imitation maps contain only the black features in the Siegfried maps, the models trained with only the imitation maps (Scenario 2) and the mixture of real and imitation maps (Scenario 3) can only be applied to the black layer of the real Siegfried map. Thus, we develop Scenario 1b. Scenario 1a serves as a comparison with 1b to investigate the impact of applying CIS before the training process. The method to extract the black layer is elaborated in Section 5.4.3.

(2) Training with imitation maps only (Scenario 2).

The four imitation map sheets that cover Zurich city are used for training.

(3) Training with mixed real and imitation maps (Scenario 3).

Two Siegfried map sheets and two imitation map sheets, which in combination cover Zurich city are used for training.

#### 5.4.3. Extracting the black layer from Siegfried map

For training Scenarios 1b, 2 and 3, the black layer of the real historical maps is required for prediction and/or training. Thus, the mean-shift clustering algorithm and threshold method, as typical CIS techniques, are applied to the Siegfried map to obtain the black layer (Chiang and Knoblock, 2009b). In mean-shift the spatial distance is set as 5 pixels and colour distance is set as 50. Figure 5.11(a) shows a small section of a Siegfried map sheet, and the result processed by mean-shift is displayed in Figure 5.11(b). Subsequently, by applying the threshold method, the pixels with RGB values smaller than or equal to (120, 120, 120) are preserved, while the other pixels are “discarded”. The values of the preserved pixels are set as (0, 0, 0), namely black pixels, while the values of other pixels are set as (255, 255, 255), namely white pixels. Figure 5.11(c) shows the result of the threshold method.

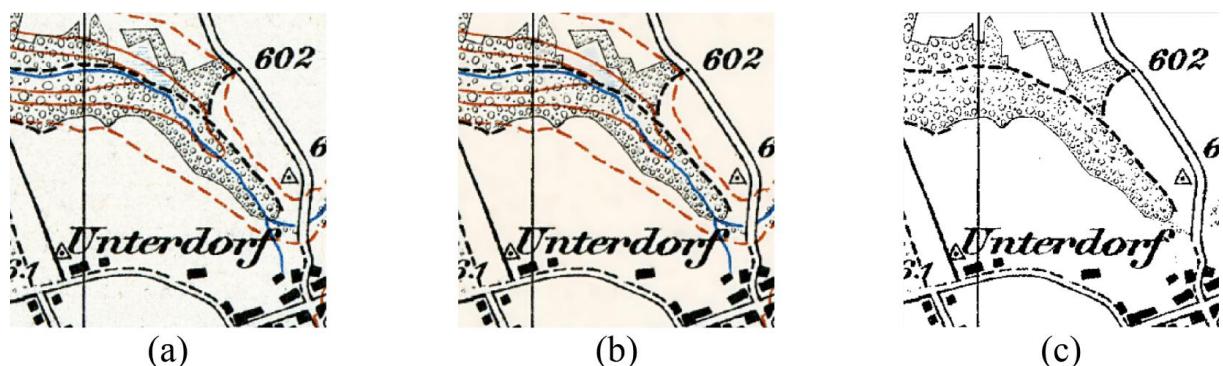


Figure 5.11. Extracting the black layer from Siegfried map, (a) the original map, (b) the result after mean-shift processing, (c) the black layer extracted by the threshold method. Geodata © Swisstopo

## 5.5. Experiment

The open-source Keras library is used to implement and train the models. Adam optimiser is used for stochastic optimisation, and the learning rate is initialised as 0.001 (Kingma and Ba, 2014; He *et al.*, 2019). Road areas and non-road areas can be quite imbalanced, especially in non-urban areas, where only narrow paths run through. Thus, we employ dice loss as the loss function (Dice, 1945; Milletari *et al.*, 2016). Each model is trained through 100 epochs, and five models are trained for each scenario to minimise the randomness. The batch size is 64. To accelerate training, the scenarios were trained on two different machines. Scenarios 1a and 1b were trained with NVIDIA GeForce RTX 2080 Ti GPU, and scenarios 2 and 3 with NVIDIA Quadro M2000 GPU. Although their technical specifications are different, in practice there are no significant differences in the results, since the model configurations and meta parameters were identical. After the training process finishes, the models are applied to four Siegfried map sheets, namely, one covering an area in the northern part of Switzerland, which contains a city and hills (sheet number TA\_017\_1940), one covering rural area in the Swiss plateau (sheet number TA\_199\_1941), one covering rural area in the periphery of the Swiss Alps (sheet number TA\_385\_1941), and the last covering small cities/towns in the eastern part of Switzerland (sheet number TA\_219\_1944). These four map sheets can represent almost all areas in Switzerland.

### 5.5.1. Road extraction

Pixel-wise road predictions indicate the probability of the pixel being a road. Pixels with a probability greater than 0.5 are regarded as roads. Then, the road areas are skeletonized using morphological operations (Jiao *et al.*, 2021). The “raster to polyline” tool in ArcGIS is used to vectorize and simplify the skeletons. Table 5.1 shows the qualitative road extraction results of the four training scenarios with 20,000 training samples. We select the results in three typical areas in the map, namely urban area, suburban area, and rural area. In the pictures of Table 5.1, blue lines present the vector road predictions, which overlay on the corresponding map images. To highlight the road lines, the underlying map images in the third to seventh rows are shown with 50% transparency.

In Result 1a and Result 1b, almost all road centerlines are well extracted, and there are rarely false positives around labels, administrative borders, etc. In Result 2, a few false positives are seen around labels, which may be attributed to some slight difference of character font and size between the imitation and real maps. In addition, for some road segments, the two parallel road lines are extracted, rather than the centerline, indicating that the models recognize the two parallel lines as paths. The reason is that the width of road lines is a bit inconsistent on the Siegfried map due to its poor quality. Although we try our best to mimic the inconsistency by varying the width of road lines, this variance may not cover all the width inconsistency. Promisingly, these two problems are solved to a large extent by mixing the imitation and real maps (Result 3). Scenario 3 represents the use case in which people only have a small amount of real training data, but a large amount of data have to be produced very quickly, so they complement the real data with the synthetic ones. In general, all the trained models in the four training scenarios are able to tackle the complexity of various features in urban, suburban, and rural areas. However, Scenarios 1a and 1b demand much time and effort to label the training data.

Table 5.1. Qualitative road extraction results of the four training scenarios with 20,000 training samples. Geodata © Swisstopo

	Urban area	Suburban area	Rural area
Original map image			
Result 1a (trained on the real Siegfried map)			
Result 1b (trained on the black layer of the real Siegfried map)			
Result 2 (trained on the imitation map)			
Result 3 (trained on the mixed real and imitation map)			
Ground truth			

### 5.5.2. Evaluation

As we have both raster road predictions and vectorized road centerlines, we employ two sets of metrics for evaluation. Accuracy and F1 score are used for the raster predictions, and completeness and correctness for the vector road centerlines (Wegner *et al.*, 2013). The metric values are “computed using the number of correctly or wrongly predicted pixels, namely true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN)” (Jiao *et al.*, 2021). The quantitative results are presented in Table 5.2. Result 1a, Result 1b, Result 2 and Results 3 have the same meanings as those in Table 5.1. In this case, we generate three-meter buffers for the extracted and the ground-truth road centerlines for computing completeness and correctness. Table 5.2 reports the mean values of accuracy, F1, completeness and correctness of the four map sheets. In addition, the average training time for each scenario is also included in Table 5.2.

Generally, with the number of training samples increasing, the performance gets more satisfying. The raster predictions achieve high metric values, especially accuracy. For vector predictions, the completeness values are very high, indicating most of the roads are extracted. Usually, Result 1b outperforms Result 1a, verifying the effectiveness of applying image processing technique (CIS in this study) before training, as the removal of coloured features (e.g., isolines, hydro features) makes Scenario 1b simpler than Scenario 1a. Notably, the results of merely using the imitation map (Result 2) are already satisfactory. By mixing the real and imitation maps, the results (Result 3) get far better, especially in terms of F1 and correctness. A higher correctness value indicates fewer false positives. For example, Figure 5.12 shows Result 3 has no false positive on the character “f”, but other results have the false positive. Figure 5.13 shows Result 3 has apparently fewer false positives along grid lines, forest borders and the label. Although Result 1a and Result 1b are already promising, they both require considerably long time (about 100h per sheet) to generate the training data. However, it is much faster to produce the imitation map. With reconstructed symbols, it takes only 2-3 h to produce an imitation map sheet using ArcGIS with the instruction in Section 5.4.1. The results of the mixed maps are very close to Result 1a and Result 1b, and the correctness values are even higher than those of Result 1a and Result 1b.

Table 5.2. Quantitative road extraction results of the four training scenarios.

Samples	Metrics	Result 1a	Result 1b	Result 2	Result 3
20,000	Accuracy	97.98%	97.98%	95.30%	97.94%
	F1 score	84.81%	84.86%	68.57%	83.76%
	Completeness	96.52%	96.58%	87.54%	93.79%
	Correctness	80.37%	80.35%	60.13%	89.02%
	Training time	30.04 hours	30.48 hours	30.20 hours	30.48 hours
10,000	Accuracy	97.60%	98.05%	95.24%	97.25%
	F1 score	81.92%	84.90%	68.51%	81.30%
	Completeness	95.88%	93.96%	88.70%	92.65%
	Correctness	75.31%	86.23%	59.87%	80.91%
	Training time	14.73 hours	14.73 hours	15.30 hours	13.55 hours
5,000	Accuracy	97.05%	97.54%	95.17%	97.42%
	F1 score	79.03%	81.78%	68.31%	80.57%
	Completeness	95.83%	95.99%	89.02%	93.83%
	Correctness	66.23%	73.41%	59.15%	75.94%
	Training time	7.58 hours	7.55 hours	7.85 hours	7.60 hours

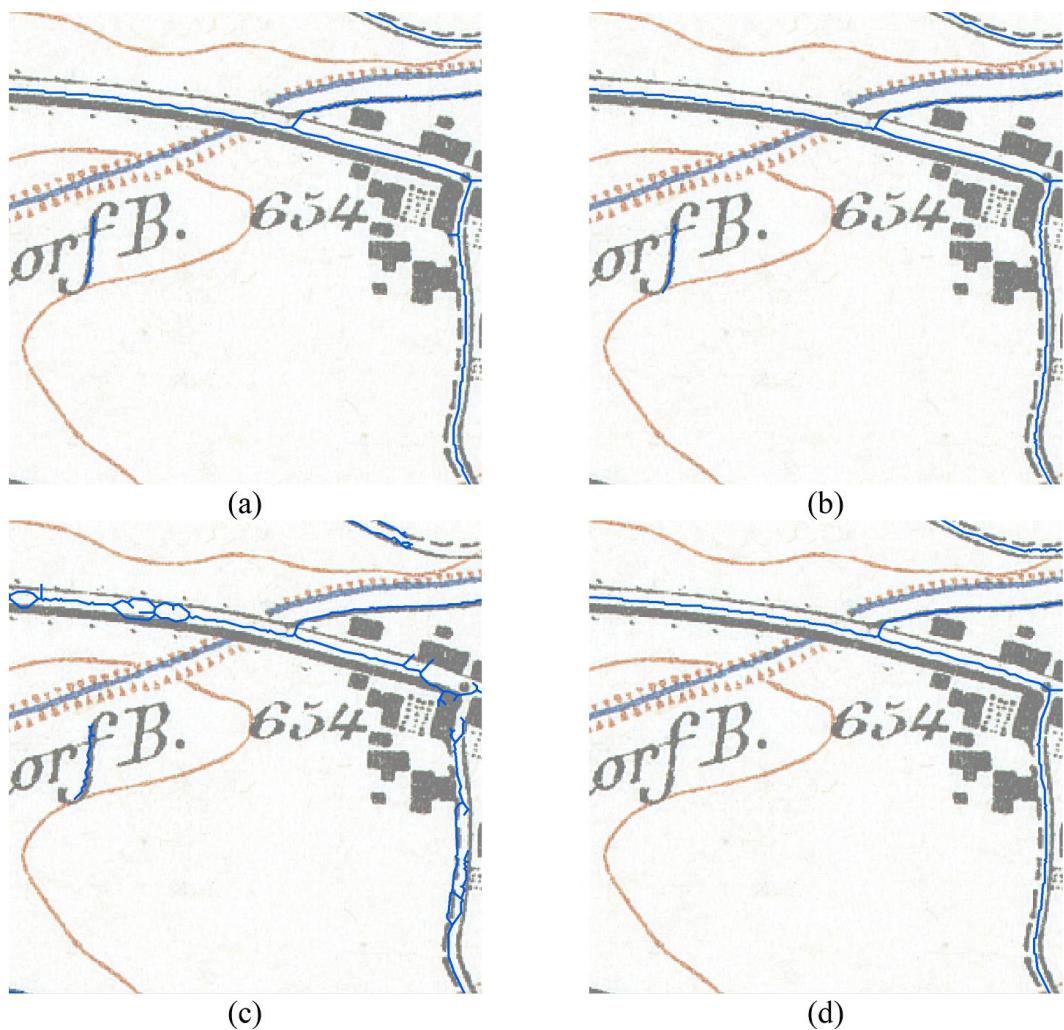


Figure 5.12. An example in which Result 3 outperforms the other results regarding false positives on the label. (a) Result 1a, (b) Result 1b, (c) Result 2, and (d) Result 3. Geodata © Swisstopo

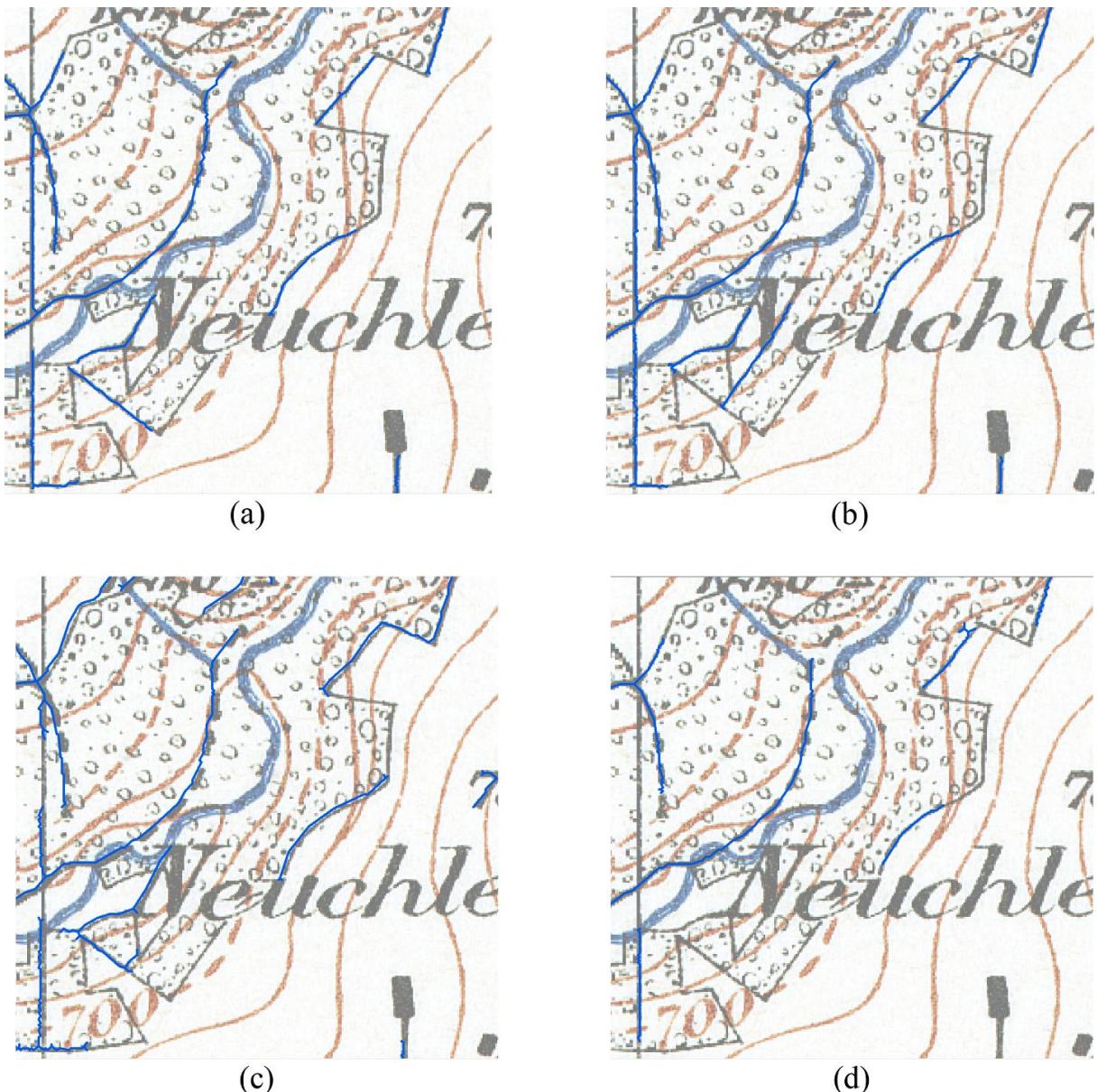


Figure 5.13. An example in which Result 3 outperforms the other results regarding false positives on the label, grid line and forest border. (a) Result 1a, (b) Result 1b, (c) Result 2, and (d) Result 3. Geodata © Swisstopo

Figure 5.14 shows the training and validation losses of the four scenarios with 20,000 samples. The losses are the mean values of the five models in each scenario. The learning process converged at around 70 epochs. In Figure 5.14(d), there is a gap (smaller than 0.01) between the training loss and the validation loss. The reason lies in that it can be more challenging for the models in Scenario 3 to generalize and to work well on the validation dataset compared with other scenarios. The models in Scenario 3 need to learn a more complex function, because its data to operate on come from two different distributions (the imitation map and the black layer of the real map) instead of only one (the real map (Scenario 1a), the black layer (Scenario 1b), and the imitation map (Scenario 2)). This means that each distribution is less represented than those in the other scenarios, essentially halving the information it can draw conclusions from. Although in Scenario 3 the validation loss looks a bit higher than the training loss, the corresponding training accuracy and the validation accuracy (in average) are as high as 97.91 % and 96.03 %, respectively. Additionally, the superiority of Result 3 shown in Table 5.2, Figure 5.12 and Figure 5.13 also verifies Scenario 3 has no overfitting issue (Ying, 2019).

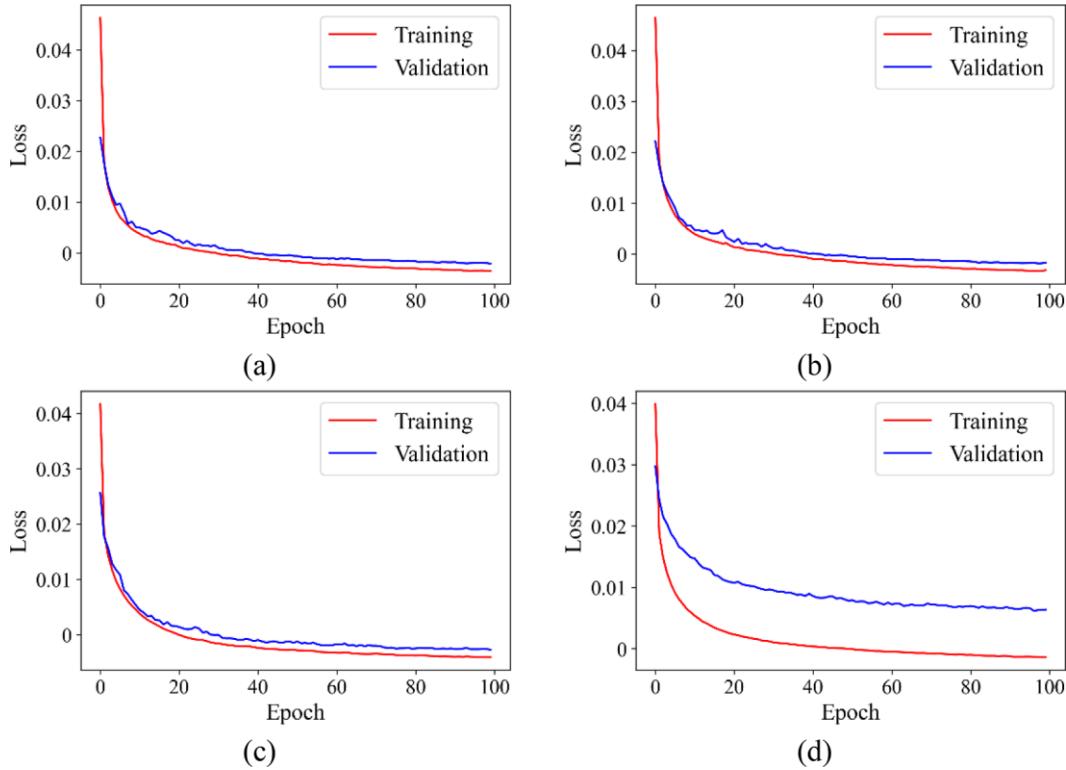


Figure 5.14. Training and validation losses of the four scenarios with 20,000 sample, (a) Scenario 1a, (b) Scenario 1b, (c) Scenario 2, and (d) Scenario 3.

## 5.6. Discussion and conclusion

To address the urgent demand for extracting roads from historical maps, this paper proposes an efficient and generalizable methodology to automatically generate road training data by symbolizing contemporary vector data and feeding them into adapted U-Net for road extraction. The whole methodology is validated with Swiss Siegfried maps. Experiments demonstrate the effectiveness of the methodology and evaluate the sensitivity of the numbers of training samples on the results.

The method of symbol reconstruction to automatically generate training data is generalizable, which only requires to reconstruct the symbols of the raster map, symbolize the corresponding vector data with these symbols and then generate the imitation map. We exemplarily apply the method to road extraction from historical maps. It should be noted that this method can be tailored and transferred to a wide range of geographical features of interest (e.g., railways, vegetation, human settlements), and to any other raster map series and possibly to other spatial data sources, such as overhead imagery. Although the study area is in Switzerland in this paper, the method is applicable to any other country.

Both using the imitation map only and mixing it with real map are feasible solutions to various situations. For instance, if shared symbolization files are accessible, these files can be used directly or with little adaptation to create the imitation map and to generate satisfactory results without real training data. Moreover, from Results 2 in Table 5.1 and Table 5.2, it is observed that increasing the number of training samples from the imitation map does not result in substantial improvement, indicating that a small amount of training samples is already enough when using merely the imitation map. Furthermore, adding the imitation map to a real training dataset can substantially improve the performance (see e.g., correctness of Result 3 for 20,000 samples). This becomes even more apparent for training with 5,000 samples. Whether to use merely the imitation map or to mix it with real training data also depends on

the desired metrics. If metrics like accuracy and completeness are required only or they are far more important than others, then a small set of imitation map data may even be enough already.

In terms of reconstructing symbols, those with homogeneous appearance (e.g., buildings shown in Figure 5.2(a) and grid lines in Figure 5.2(c)) and those consisting of regularly distributed dots/circles/strokes (e.g., gardens shown in Figure 5.2(a)) are simple to reconstruct. Moreover, point symbols represented by regular geometries are simple to create, such as the triangulation point symbolized with a triangle and a dot in the center of it in Figure 5.2(c). On the contrary, if the symbols are presented in complex patterns, it would be more challenging to reconstruct them with common GIS/cartography/graphics software. For example, the blue composite strokes represent wetlands in Siegfried map in Figure 5.15(a). The symbol consisting of clustered short horizontal strokes has random distribution patterns, which makes it challenging to reconstruct. In Figure 5.15(b), the clustered black circles and dots represent forests. The shapes, sizes and distribution of the circles are irregular, and the density of the circles and dots is higher around the forest borders. These characteristics make it challenging to reconstruct the symbols. In Figure 5.15(c), the wetland symbols superimpose the forest symbols, which would be more challenging due to the distribution of more than one set of symbols depending on each other. Most of the black circles locate where the local density of the blue strokes is quite low, so that most circles do not touch the blue strokes. Overall, for these complex symbols, the performance of symbol reconstruction for automatically generating training data still deserves investigation.

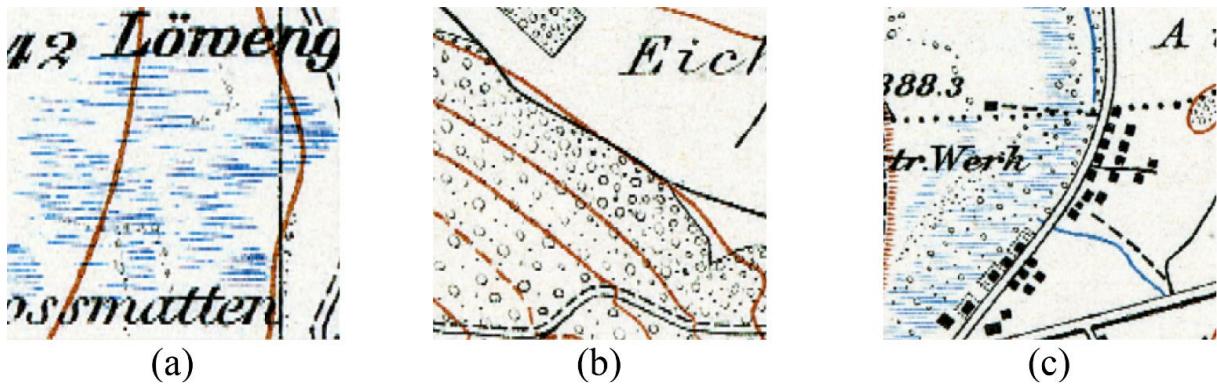


Figure 5.15. Examples of symbols that have complex patterns, (a) wetlands, (b) forests and (c) wetlands superimposing forests. Geodata © Swisstopo

Some possible improvements to the method can be seen as potential future work. First, data augmentation is used to increase the training data size and enforce CNNs to learn and identify the desired invariance of feature representations (Dosovitskiy *et al.*, 2014). Thus, it makes good sense to compare the performance of symbol reconstruction and data augmentation. Second, GANs can be employed to further improve the appearance of the imitation map to make it more resemble the real map. Third, due to limited time, only the results of 20,000, 10,000 and 5,000 training samples are reported in this paper. In the near future, more training scenarios with different numbers of samples will be experimented. Last but not least, it will be valuable to publish the constructed symbols (files), based on which people can quickly adapt the symbols to their own maps.

## Acknowledgement

We thank Ms. Monika Niederhuber for providing us the road training data of Zurich city as shown in Figure 5.3.

## Appendix

### 1. Associating the “object type” attribute with corresponding road symbol

For conciseness, the symbols shown in Figure 5.1(a)-(e) are numbered 1-5.

Railways with “object type” value “on road” are classified as tramways, which are symbolized with the symbol in Figure 5.1(f).

Table S5.1. Associating the attribute values with the corresponding symbols

<b>Attribute value</b>	<b>Road symbol number</b>
0, 1, 12	3
2, 3, 6, 8, 9, 20, 21	5
5, 10, 11	4
14, 15	2
16, 19, 22	1

### 2.Determination of road buffer sizes for ground-truth preparation

#### (1) Buffer size of the historical road data

For conciseness, the symbols shown in Figure 5.1(a)-(f) in Section 5.3.1 are numbered 1-6. Due to some uncertainty in the road width, manual measurement with the “measure” tool in ArcGIS is carried out to determine the buffer size (i.e., half of road width) of these exceptions, which is again laborious and time-consuming.

Table S5.2. Determination of road buffer sizes of the historical road data

<b>Road symbol</b>	<b>Buffer size</b>
1	~2m
2	~2m
3	~7m
4	~8m
5	~10m
6	~10m

(2) Buffer size of the contemporary road data

Table S5.3. Determination of road buffer sizes of the contemporary road data

Attribute value	Buffer size
0, 1, 12	6m
2, 3, 6, 8, 9, 20, 21	10m
5, 10, 11	8m
14, 15	3m
16, 19, 22	3m

## References

- Chen, L.C., Papandreou, G., Schroff, F. and Adam, H., 2017. Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587*.
- Chen, C.C., Knoblock, C.A. and Shahabi, C., 2008. Automatically and accurately conflating raster maps with orthoimagery. *GeoInformatica*, 12, pp.377-410. <https://doi.org/10.1007/s10707-007-0033-0>
- Chiang, Y.Y. and Knoblock, C.A., 2009a, July. A method for automatically extracting road layers from raster maps. In *2009 10th International Conference on Document Analysis and Recognition* (pp. 838-842). IEEE. <https://doi.org/10.1109/icdar.2009.274>
- Chiang, Y.Y., Knoblock, C.A., 2009b. Extracting road vector data from raster maps. In: *International Workshop on Graphics Recognition*. Springer, Berlin, Heidelberg, pp. 93–105. [https://doi.org/10.1007/978-3-642-13728-0\\_9](https://doi.org/10.1007/978-3-642-13728-0_9)
- Chiang, Y.Y., Knoblock, C.A. and Chen, C.C., 2005, November. Automatic extraction of road intersections from raster maps. In *Proceedings of the 13th annual ACM international workshop on Geographic information systems* (pp. 267-276). <https://doi.org/10.1145/1097064.1097102>
- Chiang, Y.Y., Duan, W., Leyk, S., Uhl, J.H. and Knoblock, C.A., 2020. *Using historical maps in scientific studies: Applications, challenges, and best practices* (pp. 10-25). Cham: Springer. <https://doi.org/10.1007/978-3-319-66908-3>
- Chiang, Y.Y. and Knoblock, C.A., 2013. A general approach for extracting road vector data from raster maps. *International Journal on Document Analysis and Recognition (IJDAR)*, 16, pp.55-81. <https://doi.org/10.1007/s10032-011-0177-1>
- Chiang, Y.Y., Leyk, S. and Knoblock, C.A., 2014. A survey of digital map processing techniques. *ACM Computing Surveys (CSUR)*, 47(1), pp.1-44. <https://doi.org/10.1145/2557423>
- Clevert, D.A., Unterthiner, T. and Hochreiter, S., 2015. Fast and accurate deep network learning by exponential linear units (elus). *arXiv preprint arXiv:1511.07289*.
- Dhar, D.B. and Chanda, B., 2006. Extraction and recognition of geographical features from paper maps. *International Journal of Document Analysis and Recognition (IJDAR)*, 8, pp.232-245. <https://doi.org/10.1007/s10032-005-0010-9>
- Dice, L.R., 1945. Measures of the amount of ecologic association between species. *Ecology*, 26(3), pp.297-302. <https://doi.org/10.2307/1932409>

Dosovitskiy, A., Springenberg, J.T., Riedmiller, M. and Brox, T., 2014. Discriminative unsupervised feature learning with convolutional neural networks. *Advances in neural information processing systems*, 27. <https://doi.org/10.1109/TPAMI.2015.2496141>

Duan, W., Chiang, Y.Y., Knoblock, C.A., Jain, V., Feldman, D., Uhl, J.H. and Leyk, S., 2017, November. Automatic alignment of geographic features in contemporary vector data and historical maps. In *Proceedings of the 1st workshop on artificial intelligence and deep learning for geographic knowledge discovery* (pp. 45-54). <https://doi.org/10.1145/3149808.3149816>

Duan, W., Chiang, Y.Y., Leyk, S., Uhl, J.H. and Knoblock, C.A., 2020. Automatic alignment of contemporary vector data and georeferenced historical maps using reinforcement learning. *International Journal of Geographical Information Science*, 34(4), pp.824-849. <https://doi.org/10.1080/13658816.2019.1698742>

Han, J. and Moraga, C., 1995, June. The influence of the sigmoid function parameters on the speed of backpropagation learning. In *International workshop on artificial neural networks* (pp. 195-201). Berlin, Heidelberg: Springer Berlin Heidelberg. [https://doi.org/10.1007/3-540-59497-3\\_175](https://doi.org/10.1007/3-540-59497-3_175)

He, H., Yang, D., Wang, S., Wang, S. and Li, Y., 2019. Road extraction by using atrous spatial pyramid pooling integrated encoder-decoder network and structural similarity loss. *Remote Sensing*, 11(9), p.1015. <https://doi.org/10.3390/rs11091015>

Heitzler, M., Gkonos, C., Tsorlini, A. and Hurni, L., 2018. A modular process to improve the georeferencing of the Siegfried map. *e-Perimetron*, 13(2), pp.85-100.

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

Jiao, C., Heitzler, M. and Hurni, L., 2020. Extracting Wetlands from Swiss Historical Maps with Convolutional Neural Networks. In *Automatic Vectorisation of Historical Maps. International workshop organized by the ICA Commission on Cartographic Heritage into the Digital 13 March, 2020 Budapest. Proceedings* (pp. 33-38). Department of Cartography and Geoinformatics, ELTE Eötvös Loránd University. <https://doi.org/10.21862/avhm2020.03>

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

Kingma, D.P. and Ba, J., 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Leyk, S., Boesch, R. and Weibel, R., 2005. A conceptual framework for uncertainty investigation in map-based land cover change modelling. *Transactions in GIS*, 9(3), pp.291-322. <https://doi.org/10.1111/j.1467-9671.2005.00220.x>

Li, Z., 2019, November. Generating historical maps from online maps. In *Proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems* (pp. 610-611). <https://doi.org/10.1145/3347146.3363463>

Liu, T., Xu, P. and Zhang, S., 2019. A review of recent advances in scanned topographic map processing. *Neurocomputing*, 328, pp.75-87. <https://doi.org/10.1016/j.neucom.2018.02.102>

Milletari, F., Navab, N. and Ahmadi, S.A., 2016, October. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *2016 fourth international conference on 3D vision (3DV)* (pp. 565-571). Ieee. <https://doi.org/10.1109/3dv.2016.79>

- Pezeshk, A. and Tutwiler, R.L., 2011. Automatic feature extraction and text recognition from scanned topographic maps. *IEEE Transactions on Geoscience and Remote Sensing*, 49(12), pp.5047-5063. <https://doi.org/10.1109/tgrs.2011.2157697>
- Saeedimoghaddam, M. and Stepinski, T.F., 2020. Automatic extraction of road intersection points from USGS historical map series using deep convolutional neural networks. *International Journal of Geographical Information Science*, 34(5), pp.947-968. <https://doi.org/10.1080/13658816.2019.1696968>
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. and Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1), pp.1929-1958.
- Strano, E., Nicosia, V., Latora, V., Porta, S. and Barthélemy, M., 2012. Elementary processes governing the evolution of road networks. *Scientific reports*, 2(1), p.296. <https://doi.org/10.1038/srep00296>
- Sun, T., Chen, Z., Yang, W. and Wang, Y., 2018. Stacked u-nets with multi-output for road extraction. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (pp. 202-206). <https://doi.org/10.1109/cvprw.2018.00033>
- Tsorlini, A., Iosifescu, I., Iosifescu, C. and Hurni, L., 2014. A methodological framework for analyzing digitally historical maps using data from different sources through an online interactive platform. *e-Perimetron*, 9(4), pp.153-165.
- Uhl, J.H. and Duan, W., 2021. Automating information extraction from large historical topographic map archives: new opportunities and challenges. *Handbook of Big Geospatial Data*, pp.509-522. [https://doi.org/10.1007/978-3-030-55462-0\\_20](https://doi.org/10.1007/978-3-030-55462-0_20)
- Uhl, J.H., Leyk, S., Chiang, Y.Y., Duan, W. and Knoblock, C.A., 2019. Automated extraction of human settlement patterns from historical topographic map series using weakly supervised convolutional neural networks. *IEEE Access*, 8, pp.6978-6996. <https://doi.org/10.1109/access.2019.2963213>
- Wang, J., Song, J., Chen, M. and Yang, Z., 2015. Road network extraction: A neural-dynamic framework based on deep learning and a finite state machine. *International Journal of Remote Sensing*, 36(12), pp.3144-3169. <https://doi.org/10.1080/01431161.2015.1054049>
- Wegner, J.D., Montoya-Zegarra, J.A. and Schindler, K., 2013. A higher-order CRF model for road network extraction. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1698-1705). <https://doi.org/10.1109/cvpr.2013.222>
- Ying X. An overview of overfitting and its solutions. In *Journal of physics: Conference series* 2019 Feb (Vol. 1168, p. 022022). IOP Publishing. <https://doi.org/10.1088/1742-6596/1168/2/022022>
- Yousefi-Azar, M. and Hamey, L., 2017. Text summarization using unsupervised deep learning. *Expert Systems with Applications*, 68, pp.93-105. <https://doi.org/10.1016/j.eswa.2016.10.017>
- Zhang, J., Hu, Q., Li, J. and Ai, M., 2020. Learning from GPS trajectories of floating car for CNN-based urban road extraction with high-resolution satellite imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 59(3), pp.1836-1847. <https://doi.org/10.1109/tgrs.2020.3003425>
- Zhao, P., Jia, T., Qin, K., Shan, J. and Jiao, C., 2015. Statistical analysis on the evolution of OpenStreetMap road networks in Beijing. *Physica A: Statistical Mechanics and its Applications*, 420, pp.59-72. <https://doi.org/10.1016/j.physa.2014.10.076>
- Zhou, L., Zhang, C. and Wu, M., 2018b. D-LinkNet: LinkNet with pretrained encoder and dilated convolution for high resolution satellite imagery road extraction. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops* (pp. 182-186). <https://doi.org/10.1109/cvprw.2018.00034>



## 6. A novel framework for road vectorization and classification from historical maps based on deep learning and symbol painting

---

The following chapter is a reprint of the paper “A novel framework for road vectorization and classification from historical maps based on deep learning and symbol painting” by Chenjing Jiao, Magnus Heitzler, and Lorenz Hurni, which is published by the journal *Computers, Environment and Urban Systems* <https://www.sciencedirect.com/science/article/pii/S0198971523001230>, on 14 Dec. 2023. Code and data used in this paper can be found through <https://github.com/cjiaoeth/Deep-learning-based-road-vectorisation-and-classification-with-a-painting-method>.

---

### Abstract

Road networks in the past are imperative for understanding evolution of transportation infrastructure, urban sprawl, and route planning, etc. Various approaches have been developed for road extraction from historical maps, among which deep learning techniques stand out as the most effective ones. However, little attention has been paid to investigating road vectorization and classification from historical maps. Moreover, road classification via machine learning methods usually requires large amounts of dedicated training data. To address these issues, this paper proposes a novel and comprehensive framework for road vectorization and classification on the basis of road segmentation from historical maps. First, deep learning is used to get pixel-wise raster road segmentation results, which are further skeletonized using morphological operations. Then, considering that each road class is represented with a certain symbol, a painting function is defined for each class able to paint the corresponding symbol. These painting functions are then used to draw road segments along the skeletons. Since the start and end points in each painting function are used to vectorize the segment, this method achieves vectorization and classification at the same time. Our method is validated on four Siegfried map sheets in Switzerland, and evaluated via both visual and quantitative assessments. The results indicate that the method is capable of classifying roads accurately. In particular, two evaluation metrics completeness and correctness achieve 90.69% and 72.71% respectively for road class 2 which accounts for the highest portion in the map. Moreover, the results of this method avoid the saw-toothed issue of vectorized road lines. This research is beneficial for creating complete vector road network datasets with class information to support decision-making in urban planning and transportation.

**Keywords:** Cartography, Historical maps, Road extraction, Computer-based painting, GeoAI

### 6.1. Introduction

Historical maps present geographic features in the past for large extents and various scales. They are often the only source that provides professionally surveyed geographic data (Chiang *et al.*, 2020). Road networks, as a typical anthropogenic feature, represent a physical documentation of dynamics of our civilization (Jacobson, 1940; Uhl *et al.*, 2022). Long-term spatially-explicit road data are required for the evolutional analysis of road networks (Zhao *et al.*, 2015; Casali and Heinemann, 2019), urban sprawl investigation (Nishiura and Leeruttanawisut, 2022) and spatial data conflation (Chen *et al.*, 2008; Tong *et al.*, 2014), etc. On one hand, historical maps are often the irreplaceable source of multi-temporal spatially-explicit road data that cover large spatial extents, before the emergence of born-digital maps

and overhead imagery in the late 20<sup>th</sup> century (Avcı *et al.*, 2022). On the other hand, an increasing number of systematically scanned and georeferenced historical map archives are available as public and open data, which induces the demand for automatic retrieval of information (Iosifescu *et al.*, 2016; Arcanum Maps, 2022; Swisstopo, 2022; Uhl *et al.*, 2022).

Manually vectorizing road data is tedious and time-consuming, which impedes the utilization and analysis of long-term road data, especially when regarding large extents and scales. Thus, various methods are proposed and implemented for automatic road extraction from raster maps (e.g., historical maps, current topographical maps in raster format). Some studies employ the parallel characteristic of roads on raster maps to detect road lines. For instance, in Watanabe and Oshitani (2001), road lines are identified by searching pairs of parallel line segments from urban maps. In Dhar and Chanda (2006), parallel road lines are detected from scanned topographical maps by applying the Hough transform (Ballard, 1987) to compute the orientations of lines. Nevertheless, these methods based on the parallel characteristic of roads may not be applicable to single-line road detection. Although the parallel pattern tracing algorithm reported by Chiang *et al.* (2009) and Chiang and Knoblock (2013) is capable of distinguishing a single line from double lines, it may not work for roads represented by dashed lines. In addition, Colour Image Segmentation (CIS) techniques including Mean-shift (Cheng, 1995) and K-means (McQueen, 1967) that aim at separating thematic map layers based on homogeneous colour information are employed in some other studies (Leyk and Boesch, 2010; Jiao *et al.*, 2021). For example, Chiang and Knoblock (2013) used these algorithms to cluster colours of feature layers and then asked a user to indicate road colours by manually labelling road pixels. In this way, road pixels are extracted from various map sources. Usually, CIS alone cannot completely solve the problem of road extraction. It often has to be used together with other techniques such as line tracing (Kasturi and Alemany, 1988) and morphological operation (Bovik, 2009) to get road centerlines, to fill gaps between road lines, to refine road intersections, and to remove noise (Dhar and Chanda, 2006; Chiang *et al.*, 2009; Chiang and Knoblock, 2013).

With the rise and blossom of artificial intelligence, machine learning techniques start to play an important role in road extraction and classification from historical maps. For instance, Saeedimoghaddam and Stepinski (2020) use the faster region-based deep convolutional neural network (RCNN) to extract road intersections from United States Geological Survey (USGS) historical maps. Jiao *et al.* (2022a and 2022b) employ an adapted U-Net (Ronneberger *et al.*, 2015) for road extraction from the Swiss Siegfried map. Avcı *et al.* (2022) tested U-Net++ (Zhou *et al.*, 2018) and Deeplabv3 (Chen *et al.*, 2017) for road extraction from the historical Deutsche Heereskarte. The authors also evaluated the contribution of attention blocks (Vaswani *et al.*, 2017) on road extraction by integrating the attention block to Unet++ and MA-Net (Fan *et al.*, 2020). Duan *et al.* (2020) adopt reinforcement learning (Mnih *et al.*, 2013) to automatically generate training data for road extraction from USGS historical maps. Can *et al.* (2021) developed a method to recognize road types from the Third Military Mapping Survey of Austria-Hungary historical map series using deep convolutional neural networks. Ekim *et al.* (2021) proposed a deep learning-based segmentation approach to extract multi-class roads from the Deutsche Heereskarte 1:200,000 Türkei (DHK 200 Turkey) historical maps. Both studies require time-consuming and complex steps of manually annotating and labeling road classes to generate training data. Compared with road extraction from historical maps, road extraction from overhead imagery draws far more attention (Liu *et al.*, 2019; Wei *et al.*, 2020; Li *et al.*, 2021; Chen *et al.*, 2022). However, roads on historical maps are presented with specific symbols (e.g., dashed line, solid line) instead of in their natural forms as in overhead imagery (Jiao *et al.*, 2022a). Thus, road extraction methods from overhead imagery cannot be directly applied to historical maps.

Overall, various methods have been developed to extract roads from historical maps, yet they often fall short in producing a complete road vector dataset with class information. Road class data are crucial to urban planners and policymakers in urban planning and transportation management considering that different road classes have varying traffic capacity, speed limits, and access rules (Alivand *et al.*, 2015; Zhao *et al.*, 2017). Although several studies investigated road class recognition from historical maps

using deep learning techniques (Can *et al.*, 2021; Ekim *et al.*, 2021), dedicated training data are required for road classification in terms of road class labels. Such training data are usually more labor-intensive to label with road class additionally compared with the training data for road segmentation. To the best of our knowledge, no existing contribution to the literature has attempted to classify roads into different categories from historical maps without the need of dedicated training data.

To bridge the research gap, this paper proposes a novel framework for road vectorization and classification from historical maps. This framework consists of the following processes: First, deep learning is utilized to obtain road segmentations from historical maps, on which morphological operations are applied to obtain skeletons of the road pixels. Second, the skeletons are decomposed into single curves that are continuous and follow generally similar orientations. Each single curve is approximated with a set of connected straight segments. Third, each segment is painted/symbolized with road symbols corresponding to different road classes. In this process, a loss is computed between the painted road segment and the feature in the map. The road symbol that achieves the minimum loss is adopted to label the class of the segment. In this way, the method implements vectorization and classification at the same time, and is able to produce a complete road vector dataset with class information. The framework is applied to the Swiss Siegfried map to validate its effectiveness.

## 6.2. Data

The Swiss Siegfried map is one of several main Swiss historical map series, which is published between 1872 and 1949 (Götsch, 2002). It is a comprehensive topographical map series that depicts both natural geographical features like lakes, wetlands, forests, etc. and anthropogenic features including buildings, roads, railways, toponyms, etc. The Swiss Federal Office of Topography (Swisstopo<sup>40</sup>) has scanned Siegfried map sheets into raster format, and they are georeferenced based on the map frame corner points and the coordinate grid lines (Heitzler *et al.*, 2018). The size of each scanned map sheet is  $7,000 \times 4,800$  pixels. The map sheets used in this paper are published for scale 1:25,000 and scanned with a spatial resolution of 1.25 m/pixel. Figure 6.1 shows a subsection of a Siegfried map sheet which covers a region of Zurich. Specifically for road features, there are five different classes represented by five symbols, marked by red arrows in Figure 6.1. A single dashed line represents walking path (“Fussweg”) (which is named class 1 for conciseness in this study), a single solid line represents dirt road or mule track (“Feld- oder Saumweg”) (class 2), a dashed line in combination with a solid line represents driveway without reinforcement (“Fahrweg ohne Kunstanlage”) (class 3), two parallel lines represent reinforced road 3-5 meters wide (“Kunststrasse 3-5 Meter Breite”) (class 4), a thin line together with a thicker parallel line represents a reinforced road wider than 5 meters (“Kunststrasse über 5 Meter Breite”) (class 5) (Jiao *et al.*, 2022a).

The training data available are road centerlines of Zurich city, originally produced for an internal project by the Institute of Terrestrial Ecosystems, ETH Zürich. However, the road class is not labeled in the training data. The blue lines in Figure 6.2 show the road training data, and the red polygon represents the boundary of Zurich city. Figure 6.2(a) shows the road centerlines of double-line roads, while Figure 6.2(b) the centerlines of single line roads. The orange dots represent vertices. Usually, double-line roads go through urban areas, while single line roads through suburban or rural areas.

---

<sup>40</sup> <https://www.swisstopo.admin.ch/en/home.html>

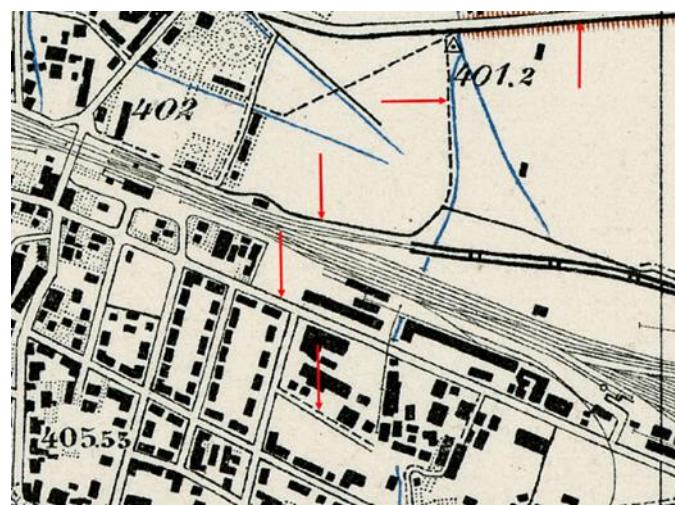


Figure 6.1. A subsection of a Siegfried map sheet covering a region in Zurich. Each red arrow points to a road class represented by a road symbol. Geodata © Swisstopo

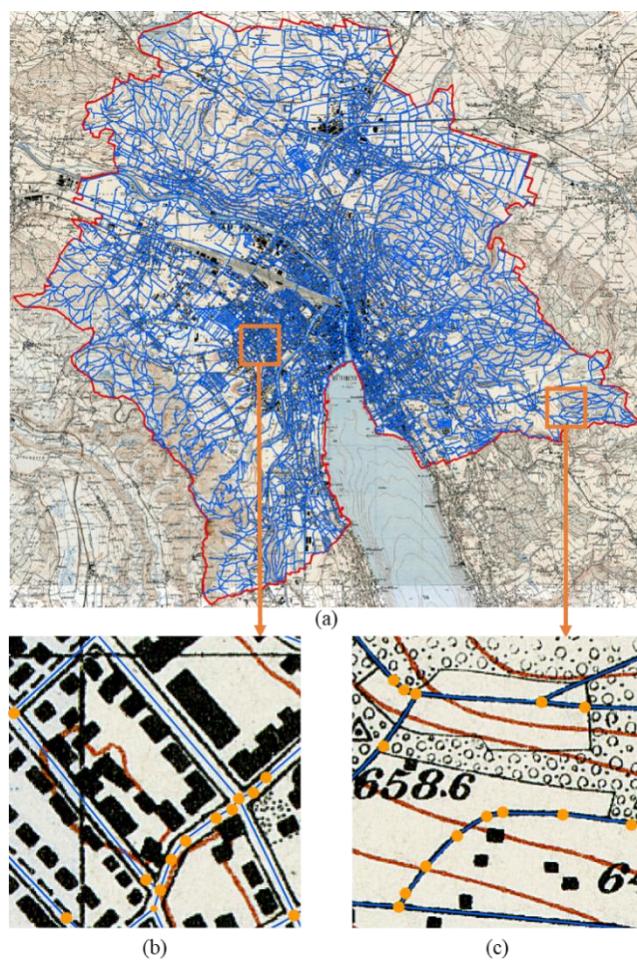


Figure 6.2. The available road training data are road centerlines (in blue). Geodata © Swisstopo

### 6.3. Methodology

The workflow of the methodology is illustrated in Figure 6.3, which is constituted of three parts, segmentation, skeleton analysis, and painting for vectorization and classification. First, the Siegfried map (a) and the road data (b) described in the previous section are used for training a U-Net (Ronneberger *et al.*, 2015) for semantic segmentation. The trained models (c) are applied to other Siegfried map sheets (d) than those used for training in order to get raster road predictions, i.e., the segmentation results (e). Secondly, morphological operations are adopted to get the skeleton (f) of the road segmentations. The raster skeletons are clipped into patches sized 128×128 pixels as shown in (g). Connected component analysis (CCA) (He *et al.*, 2017) is used to get the connected skeleton curves. As an example shown in (h), there are two connected curves (labeled with 1 and 2). Each connected curve is then decomposed into single curves based on the connectivity of each point on the curve. In (i), there are five single curves, whose start and end points are shown with yellow dots. Subsequently, the Douglas–Peucker algorithm (Douglas and Peucker, 1973) is employed to decimate each single curve to a similar curve with fewer points, which is a generalized sequence of connected segments approximating the single curve. Since each segment between a pair of points is a straight line, which can be painted with a road symbol. In (j), the red dots represent the simplification points derived by the Douglas–Peucker algorithm on each single curve. Thirdly, the five road symbols in (k) are used for painting or symbolizing each straight segment between two adjacent points. The determined road symbols are shown in (l), with the black layer of the Siegfried map in (m). The black layer refers to the layer that contains features represented with black symbols (Räth *et al.*, 2023), which is used as the target for painting, as all roads are on the black layer. In this study, we use Mean-shift to segment the black layer from the Swiss Siegfried map (Jiao *et al.*, 2022a). As illustrated in (n), all the segments are classified according to the determined road symbols. Key steps are depicted in the following sections.

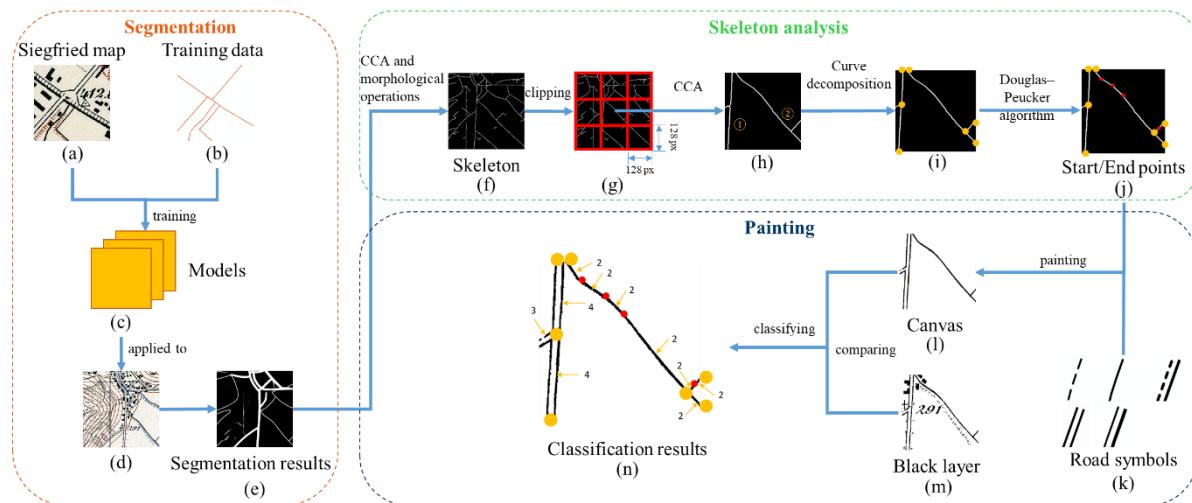


Figure 6.3. Workflow of the methodology. Geodata © Swisstopo

#### 6.3.1. Road segmentation

In this study, we utilize an adapted U-Net to extract road pixels. U-Net has been proven to be sufficient for road segmentation from historical maps in previous studies (Avci *et al.*, 2022; Jiao *et al.*, 2022). The adapted U-Net is constituted of five down-sampling steps and five up-sampling steps, which form the contractive path (or encoder) and the corresponding expansive path (or decoder). At each down-

sampling step, the size of feature maps is halved and the number of channels is doubled. Correspondingly, the up-sampling operation doubles the size of the feature maps and halves the number of channels. In this study, the number of channels in the first convolution layer is 16. At the bottleneck, the number of channels is 256. Exponential linear unit (elu) is used as the activation function (Clevert *et al.*, 2015). Moreover, dropout is used at each down-/up- sampling step with increasing dropout rates towards the bottleneck. Dropout significantly prevents overfitting by avoiding the units co-adapting too much as well as enables to train and combine many different network architectures by randomly sampling a “thinned” network consisting of all the units that survive dropout (Srivastava *et al.*, 2014; Jenny *et al.*, 2020). In the last layer, the model computes the probability of each 16-component feature vector in the produced feature maps being road. In this study, the input map tile is 128×128 pixels, while the corresponding target prediction area is 64×64 pixels, so that the network is able to learn beyond the target area to make precise predictions around the borders of the target area (Ronneberger *et al.*, 2015; Jiao *et al.*, 2022a).

The superiority of the U-Net segmentation model lies in its diametrical structure, in which the contracting path enables to capture context information and the symmetric expanding path allows for precise localization (Ronneberger *et al.*, 2015). More specifically, the contractive path generates abstract representations of the input image through deeper convolutional layers and simultaneously down-samples the input image using max pooling. The expansive path takes advantage of up-sampling layers to progressively reobtain the dimensions of the input image. The concatenation empowers the network to combine basic features (e.g., edges and corners) and complex features (e.g., buildings and roads) (Heitzler and Hurni, 2020; Jenny *et al.*, 2020; Jiao *et al.*, 2020). Especially, the large numbers of filters in the upsampling part enable the network to propagate context information to higher resolution layers, thus enabling precise localization (Ronneberger *et al.*, 2015; Buslaev *et al.*, 2018). Context information enables the network to better understand the relationships between different image features and structures.

To train a U-Net model, the input ground-truth has to be raster data. Thus, the road centerlines shown in Figure 6.2 are buffered with the road width of the corresponding road class. For example, the buffer size of road class 1 is set as three meters, as the path width is usually about six meters, and the buffer size of class 3 is set as seven meters. The historical map image is clipped into patches sized 128×128 pixels, and the patches together with the rasterized ground-truth are fed into the U-Net for training. Patches that contain roads and that do not contain roads are both sampled for training, so that the model learns both road and non-road features in order to distinguish roads from non-roads. Another advantage of this sampling strategy is that it avoids data imbalance issues, in which the occurrence of one of the classes to be distinguished is obviously rarer than the other (Tyagi and Mittal, 2020). In this case, roads on historical maps are long slender lines or areas, indicating occurrence of roads and non-roads are imbalanced, especially in non-urban areas (Jiao *et al.*, 2022a).

Furthermore, we adopt dice loss as the loss function, which is developed based on the dice coefficient (DSC) (Dice, 1945). Dice loss is proposed and designed specially for dealing with the imbalanced-data issue in image segmentation (Milletari *et al.*, 2016). DSC is one of the most common measures of region overlap. The DSC between the predictions and ground truth is calculated as in Formula 6.1.

$$DSC = \frac{2 \sum_i^H \sum_j^W p_{ij}g_{ij}}{\sum_i^H \sum_j^W p_{ij}^2 + \sum_i^H \sum_j^W g_{ij}^2} \quad 6.1$$

$i$  and  $j$  respectively denote the row number and the column number of the pixel,  $W$  and  $H$  respectively represent the width and height of the whole image, and  $p$  stands for prediction,  $g$  for ground truth. In this study, DSC is a measure of the overlap between the predictions and the ground truth in road

segmentation. If the predictions and the ground truth overlap perfectly, DSC reaches its maximum value 1. Otherwise, DSC starts to decrease, getting to its minimum value 0 if the predictions and the ground truth have no overlap at all. Thus, we use  $1 - DSC$  as the loss function to maximize the overlap between the predictions and the ground truth.

Once a well-learnt model is trained, it can be applied to other map sheets than those used for training to make predictions. The predictions are binary images with pixel values indicating whether the corresponding pixel on the map sheets is road or not, i.e., pixel value 1 indicating road while 0 non-road. Table 6.1 presents road segmentation results in urban areas, suburban areas and rural areas, from which it is qualitatively verified that the trained models are capable of recognizing roads.

Table 6.1. Road segmentation results from the Siegfried map in urban areas, sub-urban areas and rural areas.

	Urban area	Sub-urban area	Rural area
Original map image			
Segmentation result			

### 6.3.2. Road skeletonization

This section aims to skeletonize the road segmentation results. Before road skeletonization, it is necessary to remove the noise. In this study, we utilize CCA to filter out the noise in the road segmentation results. CCA is a computer vision algorithm used to detect and count the number of connected regions in a binary image. CCA assigns each region a unique label, enabling each component to be distinguished. Each region is then processed to compute a number of features of the region (e.g., gravity center, bounding box) (Bailey and Johnston, 2007). Since the road segmentation result is a binary image in our case, a region or component refers to a group of connected pixels that have the same pixel value, 0 or 1. Roads are long, slender and extensive features that should contain large amounts of connected pixels. If the size of a cluster constituted of connected road pixels is smaller than 100 pixels, which indicates the cluster can be noise such as a label or a building, the cluster will not be taken as road and will be removed. This value has been determined after experimentation and represents a heuristic measure. We are aware that different threshold values might be necessary depending on the concrete task at hand.

Then, morphological operations are used to skeletonize the road areas. A morphological operation processes an image through a shifting window, in which a local operation is carried out (Bovik, 2009).

Usually, the input of morphological operations is a binary image (e.g., the segmentation results shown in Table 6.1). Correspondingly, the output is also a binary image. Morphological skeletonization iteratively remove pixels from the object boundary while preserving the topological structure of the object (Saha *et al.*, 2016). Figure 6.4 shows an example of the road segmentation result and the corresponding skeletons of a section from the original Siegfried map. It is observed that the skeletons are saw-toothed. Moreover, road classes are not distinguished. To solve these issues, we propose and implement the method in Section 6.3.3.

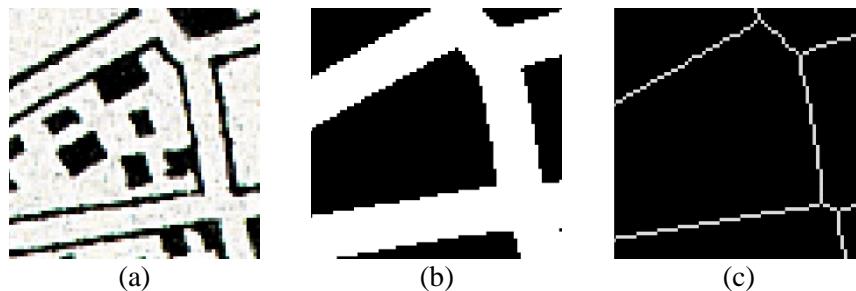


Figure 6.4. (a) A section from the original Siegfried map, (b) road segmentation result, (c) road skeletons. Geodata © Swisstopo

### 6.3.3. Road vectorization and classification

In this section, a painting-based approach is proposed to implement road vectorization and classification at the same time. In historical maps, a road class is presented with a road symbol, as depicted in section 2 and shown in Figure 6.5. Namely, road symbols can be used to determine road classes. Thus, we define a painting function for each road symbol using the OpenCV library (Bradski, 2000). As long as the parameters in a painting function have been determined, a road segment can be painted with the function. Table 6.2 expounds the parameters needed for each painting function or for each road symbol. A long road constituted of multiple segments is painted by drawing a series of connected segments, namely by applying the painting functions multiple times.

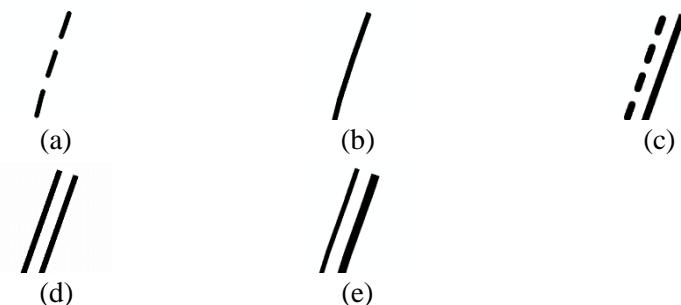


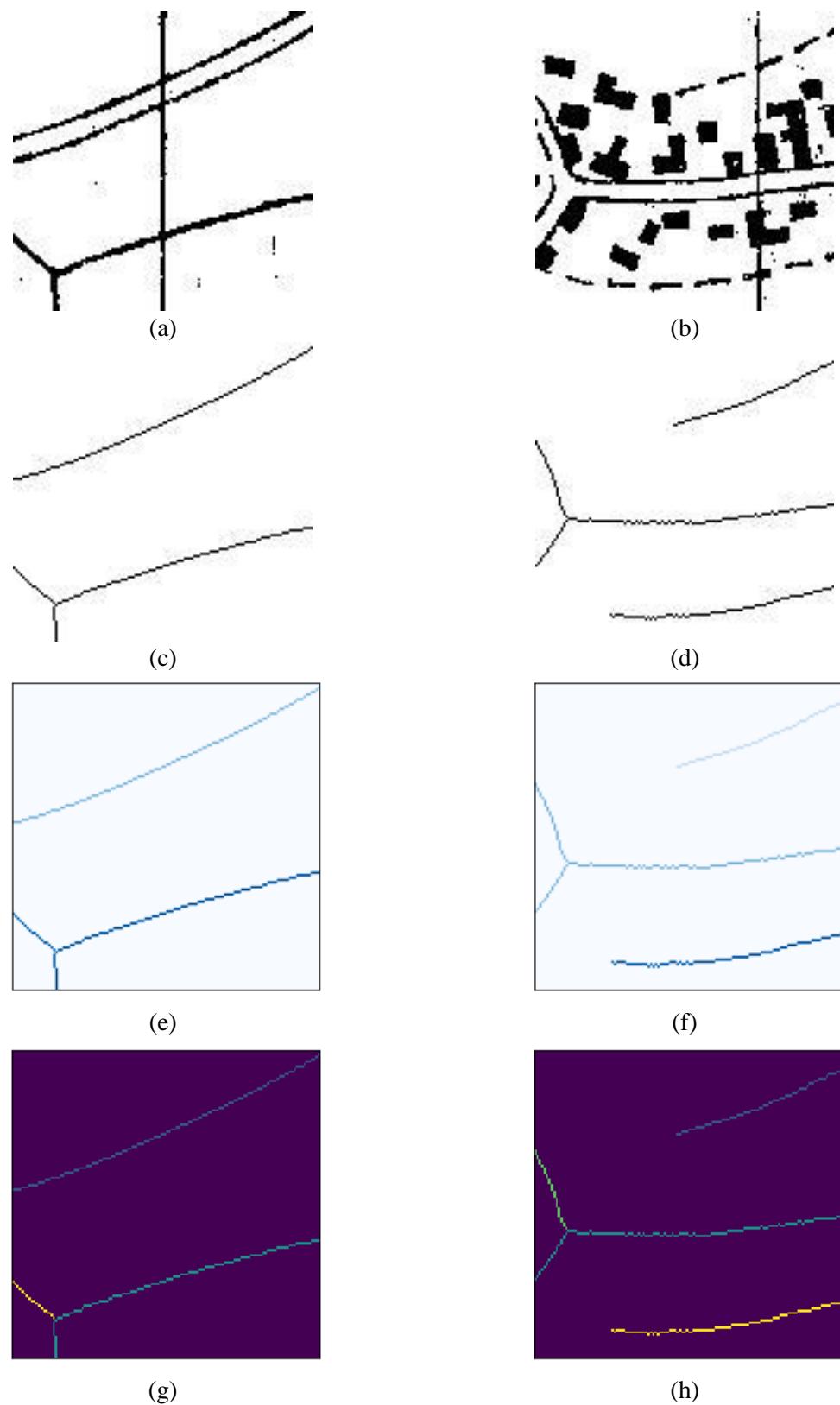
Figure 6.5. A road symbol corresponds to a road class. (a)-(e) represent road class 1-5, respectively.

Table 6.2. Parameters needed for each painting function or each road symbol.

Road class	Parameters in the painting function
1	Coordinates of start and end points, line width, length of the dash, gap between dashes
2	Coordinates of start and end points, line width
3	Coordinates of start and end points, two line widths, length of the dash of one line, gap between dashes, space between the parallel lines
4	Coordinates of start and end points, line width, space between the parallel lines
5	Coordinates of start and end points, two different line widths, space between the parallel lines

Most roads on the Siegfried map are represented by curved lines, either single or double lines. However, the road symbols are “straight”. Thus, it is necessary to find a chain of connected symbolized segments through the painting process, which best approximates the curved lines. This means that the curves must be “decomposed”, which can be achieved by splitting up the skeletons. In this study, we adopt CCA to detect distinct curves in the skeleton, in which 8-neighbour connectivity is used, i.e., a pixel has eight neighbors. Figure 6.6 shows two examples of skeleton curve decomposition in two patches. The two patches from the black layer are shown in Figure 6.6(a) and (b), the corresponding skeletons are shown in Figure 6.6 (c) and (d). Figure 6.6(e) and (f) present the distinct curves shown with different colours. However, some distinct curves contain branches pointing to multiple directions (e.g., the curve in the lower part of (e) has two branches). Thus, these curves are further decomposed by cutting the curve at the point that has a degree value higher than two. The concept of degree originates from the field of graph and network analysis. The degree of a node in a network is the number of connections or edges it has to other nodes (Dorogovtsev *et al.*, 2001; Zhao *et al.*, 2017; Pun *et al.*, 2019). In this study, we apply the concept of degree to skeletons of road networks. Each pixel/point in the skeleton is regarded as a node. Usually the degree of a node/point is two. When the node/point is located at the branch, its degree value is larger than two, which is employed to cut the distinct curves into single curves. Figure 6.6(g) and (h) show the single curves with different colours. It is seen that the distinct curve in the lower part of (e) as well as the one in the middle part of (f) are cut into two single curves.

The key of approximating a single curve is to find representative points that simplify the curve into a sequence of connected straight segments. To make use of simplification points for cartographic generalization, the Douglas–Peucker algorithm can be used. Initially, it finds the point that is farthest from the segment with the first and last points of the curve as end points. This point is kept if its distance to the segment is larger than a predefined parameter  $\epsilon$ . Then it recursively calls itself for the curve between the first point and the farthest point kept in the last iteration as well as for the curve between the last point and the farthest point, until in the curve no point farther than  $\epsilon$  from a corresponding segment can be found. In this way, the algorithm recursively divides the curve into connected straight segments by finding the simplification points (Ramer, 1972; Douglas and Peucker, 1973). This algorithm is used to find the representative simplification points on each single curve in this study. In this study, the parameter  $\epsilon$  is set as 1.5 after multiple attempts. Figure 6.6(i) and (j) show those simplification points on the single curves in (g) and (h), respectively.



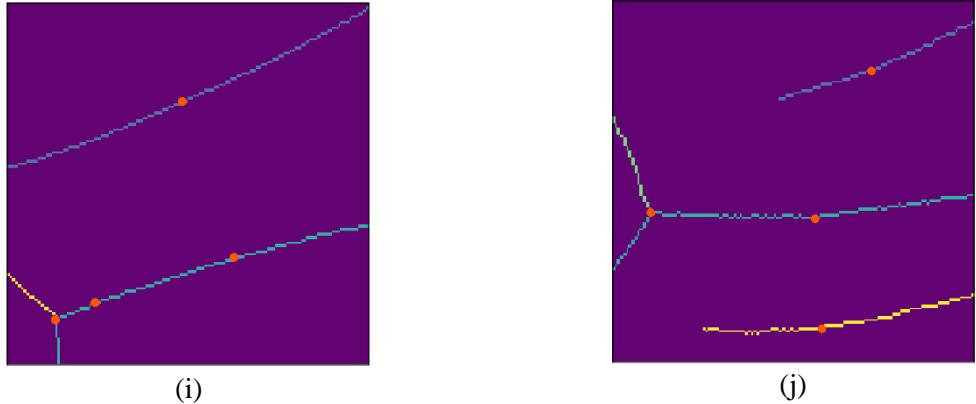


Figure 6.6. Two examples of skeleton curve decomposition. (a) and (b) present two patches from the black layer, (c) and (d) the corresponding skeletons, (e) the distinct curves of (c) decomposed by CCA represented in blue and light blue, (f) the distinct curves of (d) represented in blue, light blue and light cyan, (g) the single curves of (e) shown in blue, emerald green and yellow, (h) the single curves of (f) shown in blue, bright green, emerald green and yellow, (i) and (j) the simplification points on each single curve in (g) and (h) respectively, represented with red dots.

Once the simplification points are obtained, each road segment between a pair of simplification points can be painted with the five road symbols presented in Figure 6.5. A loss is computed between the painted/symbolized road segment and the black layer with Formula 6.2, where  $T_{ij}$  represents the value of pixel  $(i, j)$  in the target image, i.e., the black layer in this case,  $P_{ij}$  the value of pixel  $(i, j)$  in the painted canvas,  $h$  and  $w$  are the height and width of the image/canvas respectively. In Formula 6.2, the difference between normalized values of corresponding pixels in the canvas and the target is calculated, and the loss is computed as the average value of differences of all pixels. The symbol that achieves the minimum loss value is used to label the class of the segment, and the start and end points are used to vectorize the segment. Additionally, we assign priorities to road symbols in the painting process based on the frequency of that class occurring on the Siegfried map. For example, Class 2 is assigned the first priority, as its frequency is higher than all the other classes. Class 3 is assigned the fifth priority, as its frequency is only higher than tramway. When two painting functions achieve the same loss values for a road segment, the road symbol with higher priority will be assigned to the class of road segment. To accelerate the painting process, the black layer is clipped into patches sized  $128 \times 128$  pixels. Correspondingly, the canvas is sized  $128 \times 128$  pixels. Roads on the whole map sheet are painted patch by patch.

$$\text{Loss} = \frac{1}{w \cdot h} \sqrt{\sum_{j=1}^w \sum_{i=1}^h (T_{ij}/255 - P_{ij}/255)^2} \quad 6.2$$

Some painting results based on the five road symbols are shown in Figure 6.7, where the first column presents the painted roads and the second column displays the corresponding patches from the black layer. It can be observed that all the road symbols from the three patches are painted properly. One point should be noted is that the painted segment in the bottom left of Figure 6.7(c) is straight instead of curved. This is because the Douglas–Peucker algorithm is used to approximate each curve, which implies that the distance between the curve and the corresponding segment is less than the determined parameter  $\epsilon$  instead of always being 0. Even though the curve in the bottom left is painted as a straight line, its impact on road vectorization and classification is tiny. Overall, Figure 6.7 demonstrates that the painting method is capable of precisely distinguishing multiple road classes in one patch.

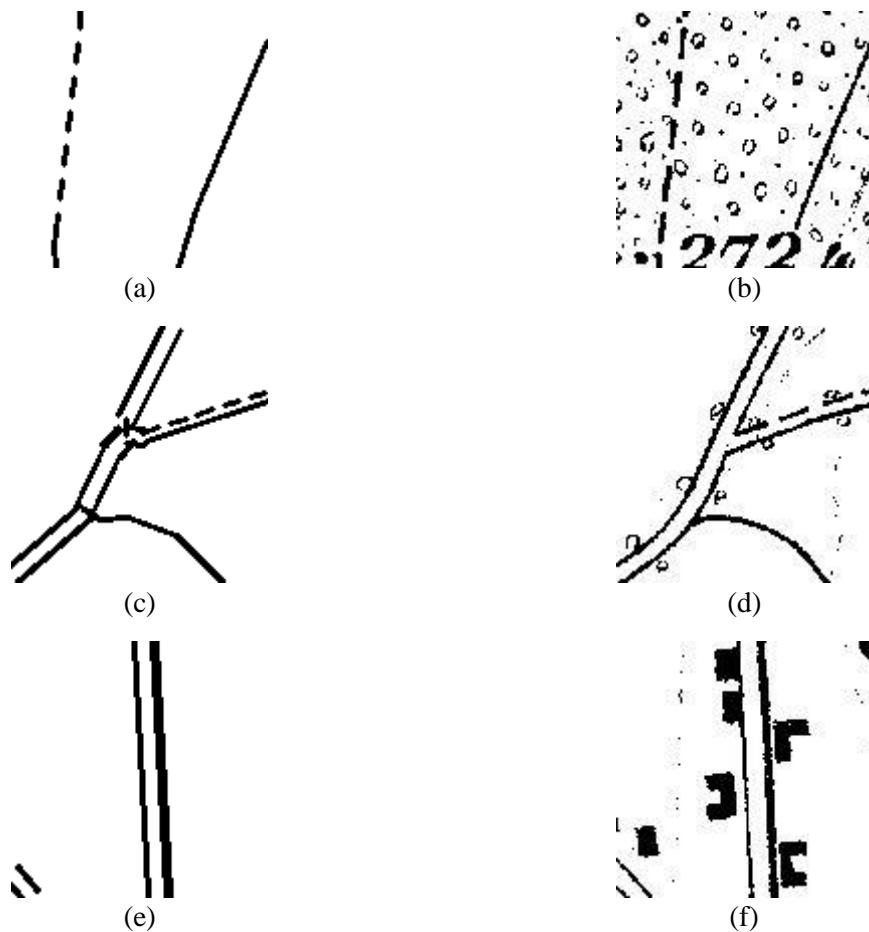


Figure 6.7. Examples of painting results. (a) shows the result of (b) including road class 1 and 2, (c) shows the result of (d) including class 1, 3 and 4, (e) shows the result of (f) including class 4 and 5.

#### 6.4. Experiment

The open-source Keras library is used to implement and train the models, and the OpenCV library is adopted to implement the painting process. In the training process, weights are initialised using the method proposed by He *et al.* (2015), which helps with convergence of very deep networks trained directly from scratch. The trained models are applied to four Siegfried map sheets that generally represent almost all areas in Switzerland, namely, one covering an area in the northern part of Switzerland, which contains a city and hills (sheet number TA\_017\_1940), one covering rural area in the Swiss plateau (sheet number TA\_199\_1941), one covering rural area in the periphery of the Swiss Alps (sheet number TA\_385\_1941), and the last covering cities/towns (sheet number TA\_151\_1940). The painting method takes only around 10 minutes to vectorize and classify one map sheet.

In this section, the painting process is illustrated, followed by results of road vectorization and classification from two aspects. First, the painting process is introduced by visualizing the loss values under different painting functions. Second, a visual assessment is conducted to validate the proposed framework. Third, we present the quantitative assessment results based on two evaluation metrics, namely completeness and correctness. In the classification results, the five road classes are visualized with the following colours. Road class 1 (single dashed line) is represented with light green lines, class 2 (single solid line) with red lines, class 3 (a dashed line in combination with a solid line) with dark green lines, class 4 (two parallel lines) with yellow lines and class 5 (a thin line together with a thicker parallel line) with orange lines, as shown in Figures 6.8 and 6.9.

#### 6.4.1. Painting process

In the painting process, a key step is to determine the optimal parameters for each painting function. The coordinates of start and end points of a segment can be obtained from the Douglas-Peucker algorithm. The other parameters for each painting function are determined manually by checking the corresponding numbers of pixels (e.g., line width, spacing between lines) in the black layers of historical maps. Table 6.3 displays the determined parameters for each painting function.

Table 6.3. The selected parameters for each painting function or each road symbol.

Road class	Road symbol	Parameters in the painting function (in pixel)
1	single dashed line	line width: 2 length of the dash: 8 gap between dashes:6
2	single solid line	line width:2
3	dashed line in combination with solid line	line widths:2 length of the dash:12 gap between dashes:8 space between the parallel lines: 6
4	two parallel lines	line widths:2 space between the parallel lines: 10
5	thin line together with thicker parallel line	thin line width:2 thick line width: 4 space between the parallel lines: 10

Based on the selected parameters for the painting functions (or for all five road classes), road segments from an image patch of the black layer are painted/symbolized. A loss value is computed by comparing the painted canvas against the image patch. Then, by comparing the loss values among the five painted results, a road segment is classified. The first column in Table 6.4 shows an image patch, the second column presents the painted canvas, the third column the difference between the image patch and the canvas, the forth column the loss values. It is observed that the symbol/painting function of a single solid line achieves the minimum loss. Thus, the road segment is classified as Class 2. In our method, the road intersections are not painted separately. The point where two vectorized road segments meet constitutes a road intersection, regardless of whether the two roads belong to the same category or not.

Table 6.4. An example of comparing the loss values of painting an image patch.

Image patch	Painted results	Difference	Loss value
			0.130
			0.124
			0.230
			0.265
			0.292

#### 6.4.2. Visual assessment

We visualize a random sample of map patches in this subsection. Figure 6.8 presents examples of painting and vectorization results in town/urban areas, where the first column shows the map patches, the second column shows the corresponding painted canvas and the third column displays the vectorized and classified road lines. In the first row, the map patch (Figure 6.8a) contains the five classes of roads and a number of buildings around them. By comparing the map patch with the vectorized roads (Figure 6.8b), it is found that all the roads are vectorized or symbolized correctly except the road intersection on the top left of the patch. By comparing the road classification result (Figure 6.8c) with the map patch, it can be observed that all the roads are classified correctly except the road intersection on the top left and the short dashed line on the bottom left. The road intersection and dashed line are misclassified into solid lines. In the second row, the map patch (Figure 6.8d) contains two road classes, i.e. a thin line together with a thicker parallel line and two parallel lines. Figure 6.8(e) shows the corresponding vectorized roads, which are consistent with the roads in the map patch. In particular, the road roundabout is symbolized as a series of connected two parallel lines. By comparing the road classification result (Figure 6.8f) with the map patch, it can be seen that the two classes are classified correctly except for a road intersection in the road turntable. In the third row, the map patch (Figure 6.8g) contains only one road class (i.e. class 4), and three roads are intersected. By comparing the map patch with the vectorized roads (Figure 6.8h), it is found that all the roads are vectorized correctly including the road intersections. The classification result is shown in Figure 6.8(i), which is correct. The visualization results demonstrate that the proposed method is capable of vectorizing and classifying roads accurately, even if the complicated situations in town/city (e.g., road roundabout, intersected roads). In the last row, the map patch (Figure 6.8j) includes three road classes, i.e. class 1, class 3 and 4. As shown in Figure 6.8(k) and 6.8(l), it can be seen that most roads are classified correctly, although the dashed line on the top is misclassified as road class 2.

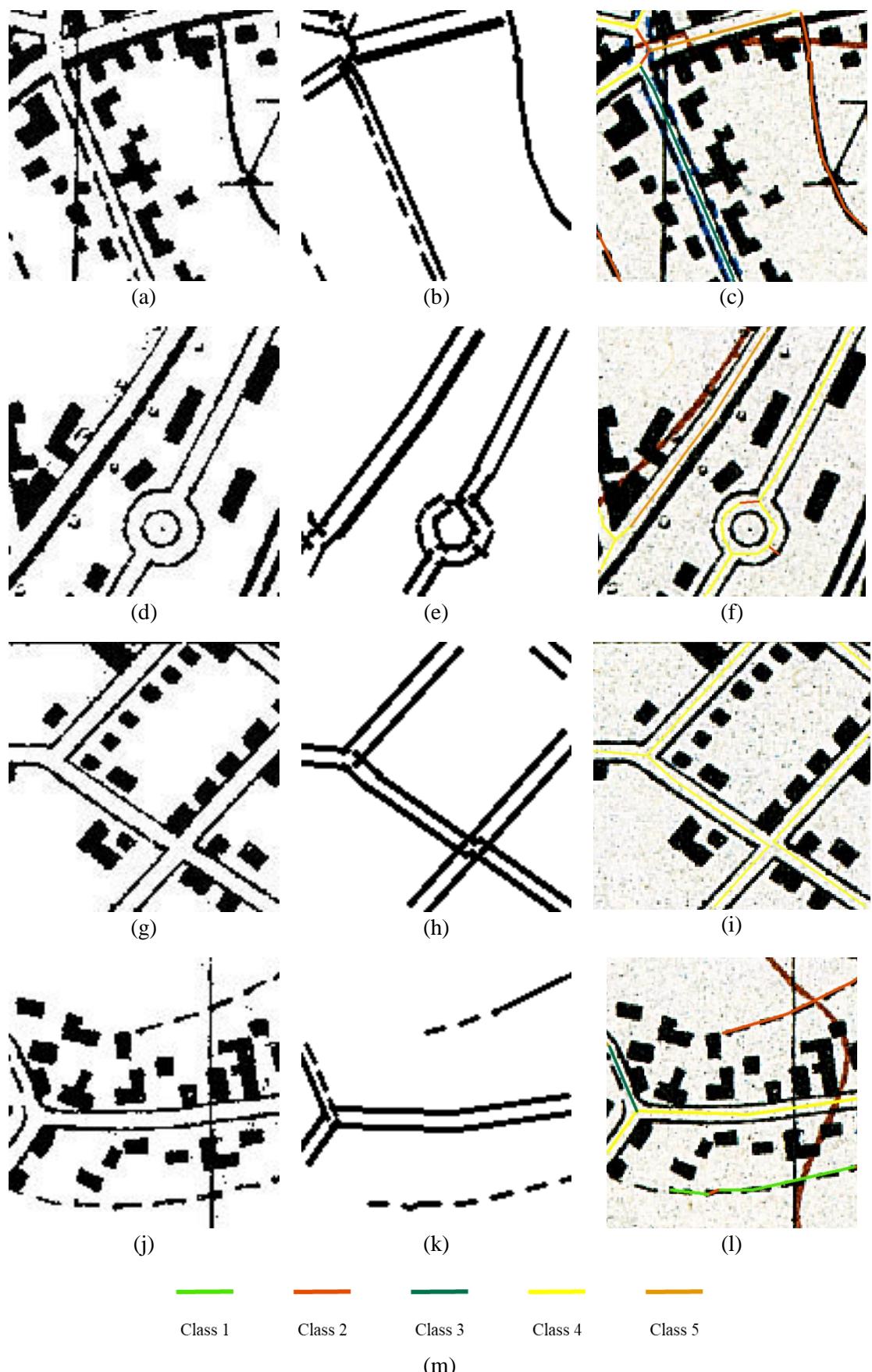
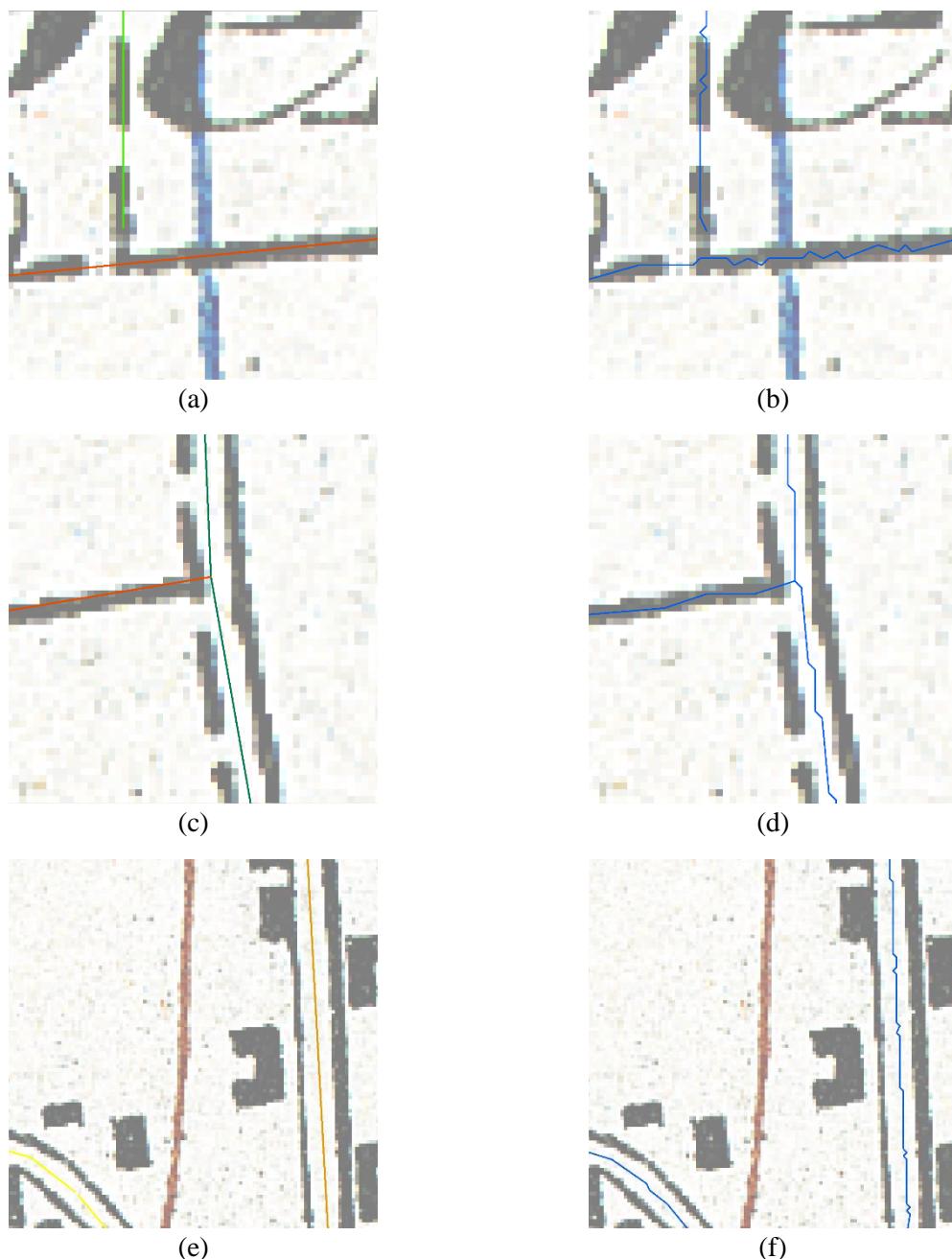


Figure 6.8. Examples of painting, vectorization and classification results in town/urban areas. (a), (d), (g) and (j) show the four selected map patches; (b), (e), (h), and (k) show the painting results respectively.

The corresponding road vectorization and classification results are presented in (c), (f), (i), and (l). (m) shows the legend for the classification results.

In order to further prove the superiority of the method, we compare the vectorization results of the painting method with the vectorized road skeletons via morphological operations at a finer scale. Figure 6.9 shows three exemplary map patches. The vectorization and classification results of the painting method are presented in the first column and the results of morphological operations in the second column. The vectorization results of morphological operations are shown in blue lines. It can be seen that the vectorized lines from morphological operations are saw-toothed, while the lines from the painting method are straight and well connected. Moreover, morphological operations do not produce classified road segments.



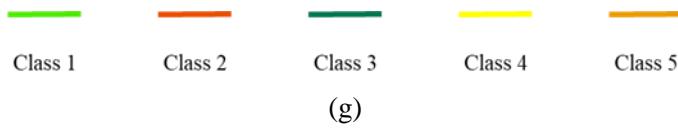


Figure 6.9. Comparison of road vectorization (and classification) results from the painting method, as shown in the first column vs. from morphological operations, as shown in the second column. (a) and (b) road class 1 and 2, (c) and (d) class 2 and 3, (e) and (f) class 4 and 5, (g) the legend.

#### 6.4.3. Quantitative assessment

In this subsection, we implement a quantitative assessment to evaluate the road vectorization and classification results. Completeness and correctness are two standard metrics for evaluating vector lines (Wiedemann, 2003), which are expressed in the following equations (Formula 6.3 and 6.4).

$$\text{Correctness} = \frac{\text{length of vectorised lines within the buffer of groundtruth}}{\text{length of vectorised lines}} \quad 6.3$$

$$\text{Completeness} = \frac{\text{length of groundtruth within the buffer of vectorised lines}}{\text{length of groundtruth vectorised lines}} \quad 6.4$$

The metric values are computed using the length of correctly or wrongly classified lines in a buffer, as denoted with Formulas 6.3 and 6.4 (Wegner *et al.*, 2013; Jiao *et al.*, 2021). In this study, we use a buffer of five meters. The average metric values of the four sheets listed beforehand are reported in Table 6.5. In the two types of single-line symbols (i.e., Class 1 and Class 2), the method yields the best classification performance in Class 2 in terms of completeness 90.69% and correctness 72.71%. A number of roads of Class 2 may be misclassified as Class 1, which explains why the correctness is not as high as completeness. In the three types of double-line symbols (i.e., Class 3, Class 4 and Class 5), the method achieves the best classification performance in Class 3 in terms of completeness 85.24% and correctness 86.25%, followed by Class 4 and Class 5. In addition, the completeness values weighted by length of each road class in the ground truth as well as the correctness values weighted by length of each road class in the extracted road lines are also displayed in Table 6.5. The weighted completeness and correctness are 77.21% and 61.34% respectively. Overall, the proposed method is capable of classifying roads except for Class 1.

Table 6.5. Average metric values of the four sheets.

Class	Completeness	Correctness
Class 1	40.97%	27.97%
Class 2	90.69%	72.71%
Class 3	85.24%	86.25%
Class 4	74.94%	69.03%
Class 5	69.85%	61.78%
Weighted	77.21%	61.34%

#### 6.4.4. Sensitivity analysis

As mentioned in Section 6.3.3, how to determine the parameter  $\epsilon$  has a significant influence on the road vectorization and classification results while applying the Douglas-Peucker algorithm to simplify/approximate each road curve into a sequence of connected straight segments. In this section, the influence of the parameter  $\epsilon$  is examined by conducting a sensitivity analysis while keeping the parameter values in painting functions fixed. Note that it may not make much sense to take into account too large or too small  $\epsilon$  values. When  $\epsilon$  is too large, the road classification accuracy will become lower. This is because the decomposed road segments can probably not be painted properly by the painting functions since they may not be “straight” enough. When  $\epsilon$  is too small, it could be time-consuming to paint all the small segments. We thus examine the impact of  $\epsilon$  values between 1 and 2 on the road classification results by taking the map sheet TA\_017\_1940 as an example.

As aforementioned, the optimal  $\epsilon$  is determined as 1.5 in this work (see Section 6.3.3). In the sensitivity analysis, we mainly compare the completeness and correctness results when  $\epsilon$  is set as 1, 1.3, 1.7, and 2 respectively. As shown in Table 6.6, it can be seen that the completeness and correctness results are very close for the five parameter settings. In addition, we further compare the road classification results with the ground truth, as shown in Figure 6.10. It can be observed that the classified roads under the five parameter settings (Figure 6.10 (b)-(f)) display highly similar spatial patterns to the ground truth (Figure 6.10 (a)). We can conclude that, varying the value of  $\epsilon$  in an appropriate range, with fixed parameter values in painting functions, has tiny influence on the final road classification results.

Table 6.6. Sensitivity analysis results for different  $\epsilon$  values.

		<b>Completeness</b>	<b>Correctness</b>
$\epsilon = 1$	Class 1	30.38	40.99
	Class 2	95.35	75.84
	Class 3	84.34	79.80
	Class 4	82.67	75.37
	Class 5	78.66	72.82
<hr/>			
$\epsilon = 1.3$	Class 1	33.05	35.82
	Class 2	94.18	76.45
	Class 3	84.46	81.12
	Class 4	82.46	77.55
	Class 5	79.44	74.07
<hr/>			
$\epsilon = 1.5$	Class 1	35.35	31.82
	Class 2	92.46	76.75
	Class 3	83.94	81.61

	Class 4	82.14	77.91
	Class 5	79.91	74.85
<hr/>			
		<b>Completeness</b>	<b>Correctness</b>
$\varepsilon = 1.7$	Class 1	36.8	27.87
	Class 2	90.56	76.87
	Class 3	83.71	81.57
	Class 4	81.31	77.93
	Class 5	79.91	74.74
<hr/>			
		<b>Completeness</b>	<b>Correctness</b>
$\varepsilon = 2$	Class 1	41.17	24.37
	Class 2	87.46	77.50
	Class 3	82.68	80.60
	Class 4	79.09	77.67
	Class 5	79.90	73.67

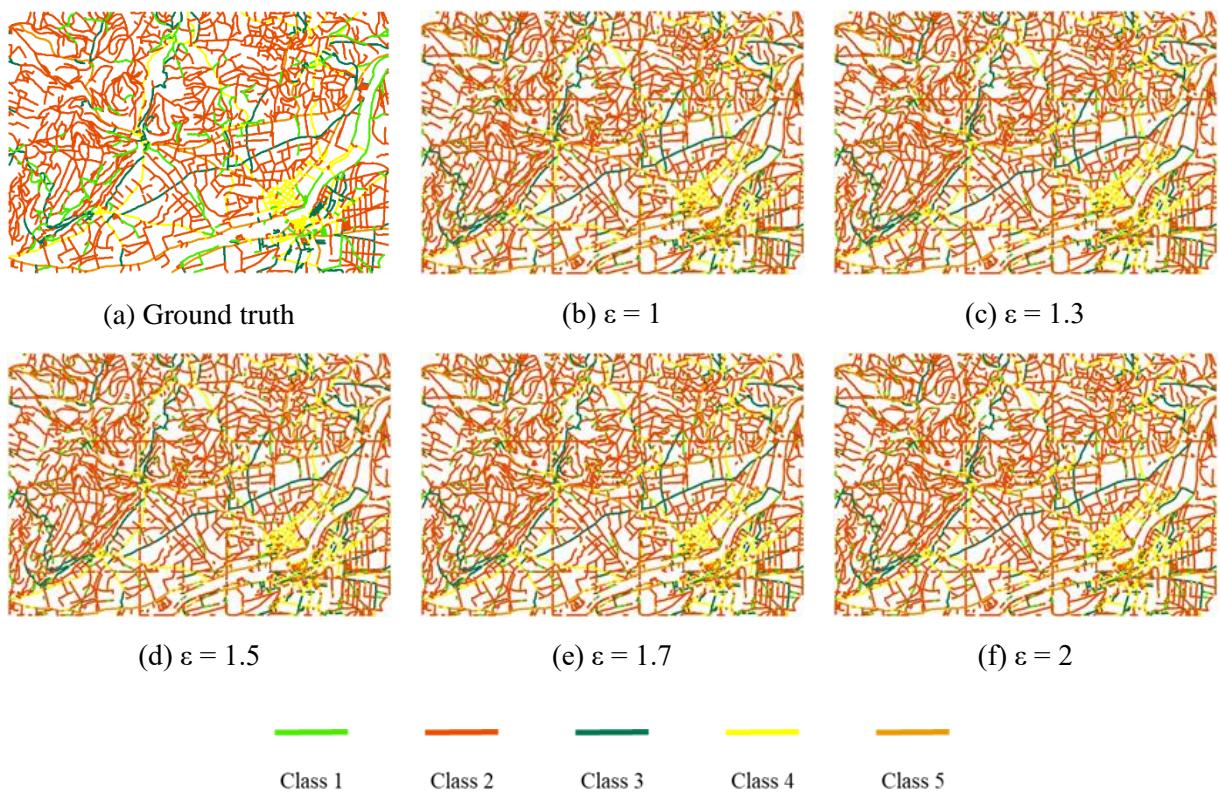


Figure 6.10. Sensitivity analysis results of road classification.

## 6.5. Discussion

The results presented in Sections 6.3 and 6.4 illustrate that the proposed framework is able to achieve accurate and robust road vectorization and classification in historical maps, with superiority to the conventional methods. Deep learning is firstly employed for getting road segmentation results, which are further used to generate road skeletons via morphological operations (Figure 6.5). The skeletons are then used as the constraints in the painting process to avoid painting in buildings and text. The visual assessment in different road scenarios (Figure 6.8) demonstrates that the proposed framework is accurate and robust in road vectorization and classification. In addition, the method is capable of solving the saw-toothed issues caused by morphological operations (Figure 6.9). Finally, the quantitative assessment results further validate the effectiveness of the proposed method in terms of correctness and completeness.

This study has several important implications with respect to road vectorization and classification as well as detection of other features from historical maps. On the one hand, compared with the road extraction results from historical maps in previous studies (Chen *et al.*, 2008; Chiang *et al.*, 2009; Jiao *et al.*, 2021), the proposed method is able to generate complete vector road datasets with class information. On the other hand, although the painting method in this study is validated on road vectorization and classification, it can be generalized to other features of historical maps. For example, painting functions of railways (line objects) and gardens (textures) can be defined based on their symbols. Then, by painting the features on a blank canvas and comparing it with the target map image, these two features can be recognized, vectorized and classified. Moreover, it should be noted that, although the proposed framework is implemented on the Swiss Siegfried maps, it can also be applicable to other historical map series, such as USGS historical maps. This is because the road symbols are often very similar among various historical map series. Even if some parameters (e.g., line width, spacing between lines) can be different among different map series, the parameter values of painting functions can be determined accordingly while applying the proposed framework to other historical maps.

The proposed framework in this study opens an avenue in road vectorization and classification from historical maps by combining deep learning with computer-based painting. There are some limitations for future research directions that we want to point out. First, road symbols or painting functions have to be defined based on some parameters. These parameters have an influence on the final road vectorization and classification results. In this study, we determine the optimal parameters manually by referencing the different road symbols in the black layer of historical maps (Table 6.3). Deep learning could be a versatile solution by training a neural network to automatically obtain parameters from symbols in the legend. Second, the proposed painting method performs not so promising for road class 1 (single dashed line), compared with the other four road symbols. The reasons behind it could be three aspects: (1) the sizes of gaps between the dashes are not consistent due to the unsatisfactory graphical quality of the historical maps, which makes it difficult to determine this parameter in the graphical painting function. (2) the start and end points of a road segment obtained from the skeletons are sometimes not perfectly the start/end points of a dash in the historical maps. (3) since single dashed lines represent walking paths that are usually winding in historical maps, they are segmented into a number of short straight segments in the painting process. This is why they are misclassified as single solid lines sometimes. Third, U-Net is used for road segmentation in the proposed framework. More modern deep learning architectures will be attempted to further improve the segmentation results since it could have a positive impact on the subsequent road vectorization and classification. Finally, tramway that occurs not so often in Swiss historical map is not taken into account in this study. Compared to the five common road classes, tramway is represented as a more complex road symbol, which deserves to be investigated in future.

## 6.6. Conclusion

This study proposes a novel and comprehensive framework for road vectorization and classification from historical maps. The framework comprises three steps: road segmentation with deep learning techniques, road skeletonization via morphological operations, and road vectorization and classification by painting. One major highlight is that the presented framework addressed the incompleteness of the extracted road data from historical maps in existing studies by developing a computer-based painting method to generate road class information without additional training data. Moreover, the proposed framework can be generalized to vectorize and classify other features (not limited to roads) from historical maps.

We implement our framework based on the Swiss Siegfried maps to showcase its effectiveness. A series of validations, including visual assessment and quantitative assessment, are conducted to justify the accuracy and robustness of the framework. The main conclusions of this study are as follows: First, the painting process intuitively shows that the painting method is capable of classifying roads by comparing loss values of the five painting functions. Second, the visual assessment results indicate the capability of the method in road vectorization and classification, even for complicated road networks in urban areas. More importantly, it avoids the creation of saw-tooth shaped lines produced by morphological operations. Third, the quantitative assessment results further demonstrate the effectiveness and accuracy of the framework. The vectorized and classified roads can be used for many potential studies, such as the analysis of road network evolution, the investigation of urban sprawl, and the design of sustainable transport infrastructures.

## References

- Arcanum Maps, 2022. Available online: <https://maps.arcanum.com/en/>. (accessed 18 November 2022).
- Avcı, C., Sertel, E. and Kabadayı, M.E., 2022. Deep learning-based road extraction from historical maps. *IEEE Geoscience and Remote Sensing Letters*, 19, pp.1-5.  
<https://doi.org/10.1109/lgrs.2022.3204817>
- Ballard, D.H., 1987. Generalizing the Hough transform to detect arbitrary shapes. Readings in computer vision. <https://doi.org/10.1016/B978-0-08-051581-6.50069-6>
- Bailey, D.G. and Johnston, C.T., 2007, December. Single pass connected components analysis. In *Proceedings of image and vision computing New Zealand* (pp. 282-287).
- Bovik, A.C. ed., 2009. *The essential guide to image processing*. Academic Press.
- Bradski, G., 2000. The openCV library. *Dr. Dobb's Journal: Software Tools for the Professional Programmer*, 25(11), pp.120-123.
- Buslaev, A., Seferbekov, S., Iglovikov, V. and Shvets, A., 2018. Fully convolutional network for automatic road extraction from satellite imagery. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops* (pp. 207-210). <https://doi.org/10.1109/cvprw.2018.00035>
- Can, Y.S., Gerrits, P.J. and Kabadayı, M.E., 2021. Automatic detection of road types from the third military mapping survey of Austria-Hungary historical map series with deep convolutional neural networks. *IEEE Access*, 9, pp.62847-62856. <https://doi.org/10.1109/access.2021.3074897>

Casali, Y. and Heinimann, H.R., 2019. A topological analysis of growth in the Zurich road network. *Computers, Environment and Urban Systems*, 75, pp.244-253.  
<https://doi.org/10.1016/j.compenvurbsys.2019.01.010>

Chen, C.C., Knoblock, C.A. and Shahabi, C., 2008. Automatically and accurately conflating raster maps with orthoimagery. *GeoInformatica*, 12(3), pp.377-410. <https://doi.org/10.1007/s10707-007-0033-0>

Chen, L.C., Papandreou, G., Schroff, F. and Adam, H., 2017. Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587*.

Chen, Z., Deng, L., Luo, Y., Li, D., Junior, J.M., Gonçalves, W.N., Nurunnabi, A.A.M., Li, J., Wang, C. and Li, D., 2022. Road extraction in remote sensing data: A survey. *International Journal of Applied Earth Observation and Geoinformation*, 112, p.102833. <https://doi.org/10.1016/j.jag.2022.102833>

Cheng, Y., 1995. Mean shift, mode seeking, and clustering. *IEEE transactions on pattern analysis and machine intelligence*, 17(8), pp.790-799. <https://doi.org/10.1109/34.400568>

Chiang, Y.Y., Duan, W., Leyk, S., Uhl, J.H. and Knoblock, C.A., 2020. *Using historical maps in scientific studies: Applications, challenges, and best practices*. Cham: Springer.  
<https://doi.org/10.1007/978-3-319-66908-3>

Chiang, Y.Y. and Knoblock, C.A., 2009, July. A method for automatically extracting road layers from raster maps. In *2009 10th International Conference on Document Analysis and Recognition* (pp. 838-842). IEEE. <https://doi.org/10.1109/icdar.2009.274>

Chiang, Y.Y. and Knoblock, C.A., 2010. Extracting road vector data from raster maps. In *Graphics Recognition. Achievements, Challenges, and Evolution: 8<sup>th</sup> International Workshop, GREC 2009, La Rochelle, France, July 22-23, 2009. Selected Papers 8* (pp. 93-105). Springer Berlin Heidelberg. [https://doi.org/10.1007/978-3-642-13728-0\\_9](https://doi.org/10.1007/978-3-642-13728-0_9)

Chiang, Y.Y., Knoblock, C.A., Shahabi, C. and Chen, C.C., 2009. Automatic and accurate extraction of road intersections from raster maps. *GeoInformatica*, 13(2), pp.121-157.  
<https://doi.org/10.1007/s10707-008-0046-3>

Chiang, Y.Y. and 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), pp.55-81.  
<https://doi.org/10.1007/s10032-011-0177-1>

Clevert, D.A., Unterthiner, T. and Hochreiter, S., 2015. Fast and accurate deep network learning by exponential linear units (elus). *arXiv preprint arXiv:1511.07289*.

Dhar, D.B. and Chanda, B., 2006. Extraction and recognition of geographical features from paper maps. *International Journal of Document Analysis and Recognition (IJDAR)*, 8(4), pp.232-245. <https://doi.org/10.1007/s10032-005-0010-9>

Dice, L.R., 1945. Measures of the amount of ecologic association between species. *Ecology*, 26(3), pp.297-302. <https://doi.org/10.2307/1932409>

Dorogovtsev, S.N., Mendes, J.F. and Samukhin, A.N., 2001. Size-dependent degree distribution of a scale-free growing network. *Physical Review E*, 63(6), p.062101.  
<https://doi.org/10.1103/physreve.63.062101>

Douglas, D.H. and Peucker, T.K., 1973. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. *Cartographica: the international journal for geographic information and geovisualization*, 10(2), pp.112-122. <https://doi.org/10.3138/fm57-6770-u75u-7727>

- Duan, W., Chiang, Y.Y., Leyk, S., Uhl, J.H. and Knoblock, C.A., 2020. Automatic alignment of contemporary vector data and georeferenced historical maps using reinforcement learning. *International Journal of Geographical Information Science*, 34(4), pp.824-849.  
<https://doi.org/10.1080/13658816.2019.1698742>
- Ekim, B., Sertel, E. and 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), p.492. <https://doi.org/10.3390/ijgi10080492>
- Fan, T., Wang, G., Li, Y. and Wang, H., 2020. Ma-net: A multi-scale attention network for liver and tumor segmentation. *IEEE Access*, 8, pp.179656-179665. <https://doi.org/10.1109/access.2020.3025372>
- Goodfellow, I., Bengio, Y. and Courville, A., 2016. *Deep learning*. MIT press.
- Götsch, C., 2002. *Siegfried-und Landeskarten: geschrieben für Sammler aus Freude an alten Karten*. Selbstverl.
- He, L., Ren, X., Gao, Q., Zhao, X., Yao, B. and Chao, Y., 2017. The connected-component labeling problem: A review of state-of-the-art algorithms. *Pattern Recognition*, 70, pp.25-43.  
<https://doi.org/10.1016/j.patcog.2017.04.018>
- He, K., Zhang, X., Ren, S. and Sun, J., 2015. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision* (pp. 1026-1034). <https://doi.org/10.1109/iccv.2015.123>
- Heitzler, M., Gkonos, C., Tsorlini, A. and Hurni, L., 2018. A modular process to improve the georeferencing of the Siegfried map. *e-Perimetron*, 13(2), pp.96-100.
- Heitzler, M. and Hurni, L., 2020. Cartographic reconstruction of building footprints from historical maps: A study on the Swiss Siegfried map. *Transactions in GIS*, 24(2), pp.442-461.  
<https://doi.org/10.1111/tgis.12610>
- Iosifescu, I., Tsorlini, A. and Hurni, L., 2016. Towards a comprehensive methodology for automatic vectorization of raster historical maps. *e-Perimetron*, 11(2), pp.57-76.
- Jacobson, H.R., 1940. *A history of roads from ancient times to the motor age* (Doctoral dissertation, Georgia Institute of Technology).
- Jenny, B., Heitzler, M., Singh, D., Farmakis-Serebryakova, M., Liu, J.C. and Hurni, L., 2020. Cartographic relief shading with neural networks. *IEEE Transactions on Visualization and Computer graphics*, 27(2), pp.1225-1235. <https://doi.org/10.1109/tvcg.2020.3030456>
- Jiao, C., Heitzler, M. and Hurni, L., 2020. Extracting Wetlands from Swiss Historical Maps with Convolutional Neural Networks. In *Automatic Vectorisation of Historical Maps. International workshop organized by the ICA Commission on Cartographic Heritage into the Digital 13 March, 2020 Budapest. Proceedings* (pp. 33-38). Department of Cartography and Geoinformatics, ELTE Eötvös Loránd University. <https://doi.org/10.21862/avhm2020.03>
- Jiao, C., Heitzler, M. and Hurni, L., 2021. A survey of road feature extraction methods from raster maps. *Transactions in GIS*, 25(6), pp.2734-2763. <https://doi.org/10.1111/tgis.12812>
- Jiao, C., Heitzler, M. and 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, p.102980. <https://doi.org/10.1016/j.jag.2022.102980>
- Jiao, C., Heitzler, M. and 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 and Spatial Information Sciences*, 2, pp.423-429. <https://doi.org/10.5194/isprs-annals-v-2-2022-423-2022>

Kasturi, R. and Alemany, J., 1988. Information extraction from images of paper-based maps. *IEEE Transactions on Software Engineering*, 14(5), pp.671-675. <https://doi.org/10.1109/32.6145>

Kato, H., Beker, D., Morariu, M., Ando, T., Matsuoka, T., Kehl, W. and Gaidon, A., 2020. Differentiable rendering: A survey. *arXiv preprint arXiv:2006.12057*.

Leyk, S. and Boesch, R., 2010. Colors of the past: color image segmentation in historical topographic maps based on homogeneity. *GeoInformatica*, 14(1), pp.1-21. <https://doi.org/10.1007/s10707-008-0074-z>

Li, J., Liu, Y., Zhang, Y. and Zhang, Y., 2021. Cascaded attention DenseUNet (CADUNet) for road extraction from very-high-resolution images. *ISPRS International Journal of Geo-Information*, 10(5), p.329. <https://doi.org/10.3390/ijgi10050329>

Li, S., Liao, C., Ding, Y., Hu, H., Jia, Y., Chen, M., Xu, B., Ge, X., Liu, T. and Wu, D., 2022. Cascaded residual attention enhanced road extraction from remote sensing images. *ISPRS International Journal of Geo-Information*, 11(1), p.9. <https://doi.org/10.3390/ijgi11010009>

Library of Congress. (2020). Sanborn maps. Available online: <https://www.loc.gov/collections/sanborn-maps/> (accessed 01 January 2020).

Liu, R., Miao, Q., Song, J., Quan, Y., Li, Y., Xu, P. and Dai, J., 2019. Multiscale road centerlines extraction from high-resolution aerial imagery. *Neurocomputing*, 329, pp.384-396. <https://doi.org/10.1016/j.neucom.2018.10.036>

McQueen, J.B., 1967. Some methods of classification and analysis of multivariate observations. In *Proc. of 5th Berkeley Symposium on Math. Stat. and Prob.* (pp. 281-297).

Milletari, F., Navab, N. and Ahmadi, S.A., 2016, October. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *2016 fourth international conference on 3D vision (3DV)* (pp. 565-571). IEEE. <https://doi.org/10.1109/3dv.2016.79>

Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D. and Riedmiller, M., 2013. Playing atari with deep reinforcement learning. *arXiv preprint arXiv:1312.5602*.

Nishiura, S. and Leeruttanawisut, K., 2022. Evolution of subcenter structure in Bangkok metropolitan development from 1988 to 2018. *Applied Geography*, 145, p.102715. <https://doi.org/10.1016/j.apgeog.2022.102715>

Pezeshk, A. and Tutwiler, R.L., 2011. Automatic feature extraction and text recognition from scanned topographic maps. *IEEE Transactions on Geoscience and Remote Sensing*, 49(12), pp.5047-5063. <https://doi.org/10.1109/tgrs.2011.2157697>

Pun, L., Zhao, P. and Liu, X., 2019. A multiple regression approach for traffic flow estimation. *IEEE access*, 7, pp.35998-36009. <https://doi.org/10.1109/access.2019.2904645>

Ramer, U., 1972. An iterative procedure for the polygonal approximation of plane curves. *Computer graphics and image processing*, 1(3), pp.244-256. [https://doi.org/10.1016/s0146-664x\(72\)80017-0](https://doi.org/10.1016/s0146-664x(72)80017-0)

Räth, Y.M., Grêt-Regamey, A., Jiao, C., Wu, S. and van Strien, M.J., 2023. Settlement relationships and their morphological homogeneity across time and scale. *Scientific Reports*, 13(1), p.11248. <https://doi.org/10.1038/s41598-023-38338-9>

Ronneberger, O., Fischer, P. and Brox, T., 2015, October. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention* (pp. 234-241). Springer, Cham. [https://doi.org/10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28)

- Saeedimoghaddam, M. and Stepinski, T.F., 2020. Automatic extraction of road intersection points from USGS historical map series using deep convolutional neural networks. *International Journal of Geographical Information Science*, 34(5), pp.947-968.  
<https://doi.org/10.1080/13658816.2019.1696968>
- Saha, P.K., Borgefors, G. and di Baja, G.S., 2016. A survey on skeletonization algorithms and their applications. *Pattern recognition letters*, 76, pp.3-12. <https://doi.org/10.1016/j.patrec.2015.04.006>
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. and Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1), pp.1929-1958.
- Swisstopo, 2022. A journey through time - Maps. Available online:  
<https://www.swisstopo.admin.ch/en/maps-data-online/maps-geodata-online/journey-through-time.html>.  
 (accessed on 18 Nov. 2022).
- Tewari, A., Thies, J., Mildenhall, B., Srinivasan, P., Tretschk, E., Yifan, W., Lassner, C., Sitzmann, V., Martin-Brualla, R., Lombardi, S. and Simon, T., 2022, May. Advances in neural rendering. In *Computer Graphics Forum* (Vol. 41, No. 2, pp. 703-735). <https://doi.org/10.1111/cgf.14507>
- Tong, X., Liang, D. and Jin, Y., 2014. A linear road object matching method for conflation based on optimization and logistic regression. *International Journal of Geographical Information Science*, 28(4), pp.824-846. <https://doi.org/10.1080/13658816.2013.876501>
- Tyagi, S. and Mittal, S., 2020. Sampling approaches for imbalanced data classification problem in machine learning. In *Proceedings of ICRIC 2019: Recent Innovations in Computing* (pp. 209-221). Springer International Publishing. [https://doi.org/10.1007/978-3-030-29407-6\\_17](https://doi.org/10.1007/978-3-030-29407-6_17)
- Uhl, J.H., Leyk, S., Chiang, Y.Y. and Knoblock, C.A., 2022. Towards the automated large-scale reconstruction of past road networks from historical maps. *Computers, environment and urban systems*, 94, p.101794. <https://doi.org/10.1016/j.compenvurbsys.2022.101794>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł. and Polosukhin, I., 2017. Attention is all you need. *Advances in neural information processing systems*, 30.
- Watanabe, T. and Oshitani, T., 2001, September. Parallel recognition of roads from urban maps on generation/verification paradigm of hypotheses. In *Proceedings of Sixth International Conference on Document Analysis and Recognition* (pp. 1225-1234). IEEE. <https://doi.org/10.1109/icdar.2001.953978>
- Wiedemann, C., 2003. External evaluation of road networks. *International Archives of Photogrammetry Remote Sensing and Spatial Information Sciences*, 34(3/W8), pp.93-98.
- Wegner, J.D., Montoya-Zegarra, J.A. and Schindler, K., 2013. A higher-order CRF model for road network extraction. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1698-1705). <https://doi.org/10.1109/cvpr.2013.222>
- Wei, Y., Zhang, K. and Ji, S., 2020. Simultaneous road surface and centerline extraction from large-scale remote sensing images using CNN-based segmentation and tracing. *IEEE Transactions on Geoscience and Remote Sensing*, 58(12), pp.8919-8931. <https://doi.org/10.1109/tgrs.2020.2991733>
- Xu, Y., Chen, H., Du, C. and Li, J., 2021. MSACon: Mining spatial attention-based contextual information for road extraction. *IEEE Transactions on Geoscience and Remote Sensing*, 60, pp.1-17. <https://doi.org/10.1109/tgrs.2021.3073923>
- Zhao, P., Jia, T., Qin, K., Shan, J. and Jiao, C., 2015. Statistical analysis on the evolution of OpenStreetMap road networks in Beijing. *Physica A: Statistical Mechanics and its Applications*, 420, pp.59-72. <https://doi.org/10.1016/j.physa.2014.10.076>

Zhao, S., Zhao, P. and Cui, Y., 2017. A network centrality measure framework for analyzing urban traffic flow: A case study of Wuhan, China. *Physica A: Statistical Mechanics and its Applications*, 478, pp.143-157. <https://doi.org/10.1016/j.physa.2017.02.069>

Zhou, Z., Rahman Siddiquee, M.M., Tajbakhsh, N. and Liang, J., 2018. Unet++: A nested u-net architecture for medical image segmentation. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4<sup>th</sup> International Workshop, DLMIA 2018, and 8<sup>th</sup> International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings* 4 (pp. 3-11). Springer International Publishing. [https://doi.org/10.1007/978-3-030-00889-5\\_1](https://doi.org/10.1007/978-3-030-00889-5_1)

## 7. Conclusions

### 7.1. Synthesis and Contributions

This thesis investigates the challenge of incorporating cartographic method knowledge into deep learning to improve feature extraction results from historical maps. Considering the broad use of road network data in various scientific domains, the cartographic element “road” is taken as the target feature. Research gaps are identified through a comprehensive review of road extraction methods mainly from raster maps. One key requirement for implementing supervised deep learning is an adequate amount of training data. Thus, two methods, i.e., novel data augmentation and symbol reconstruction are proposed and implemented to enhance the training dataset and to automatically create training data, respectively. Next, a painting method is developed to vectorize and classify raster extraction results. Qualitative and quantitative evaluations verify the effectiveness of the methodologies.

Four research questions are proposed in Chapter 1 and are investigated and solved in the following chapters. **Research Question 1** is aiming at reviewing existing road extraction methods from raster maps and overhead imagery, based on which research gaps are found. **Research Question 2** is targeted at employing cartographic method knowledge to augment training data for deep learning-based road extraction from historical maps. **Research Question 3** explores how cartographic method knowledge can be used to automatically create training data for deep learning-based road extraction from historical maps. **Research Question 4** investigates how cartographic method knowledge can help improve deep learning-based road vectorization and classification results.

To address **Research Question 1**, road extraction methods from raster maps of the past 35 years as well as from overhead imagery since 2014 are reviewed. Techniques in methods from raster maps are categorized into “line extraction, image filtering, and CIS”. Evolution of these methods is analyzed, which indicates future directions. For example, despite of the success of deep learning in image processing, its use to road extraction from historical maps is rarely seen. In the few studies that employ machine/deep learning for road extraction from historical maps, no additional cartographic method knowledge is used. Additionally, the performance of supervised deep learning relies heavily on the amount of training data. However, it is tedious and time-consuming to manually label the data. Thus, how to enhance a training dataset or even how to automatically generate training data remains unsolved. Moreover, usually road extraction results are in raster format, and vectorization is taken as “a following separate step” (Jiao *et al.*, 2021). The tasks of developing a complete framework to generate raster road predictions, vectorized road lines and road classifications are not yet tackled. Fortunately, some of the reviewed studies provide inspiration for the task of deep learning-based road extraction, i.e., integrating geo-spatial data of different sources (Wang *et al.*, 2015; Duan *et al.*, 2020).

**Research Question 2** is solved through implementing a U-Net for road extraction from the Swiss Siegfried map and developing a novel data augmentation method. In this method, only target features (roads in this use case) are rotated or flipped instead of the whole training patch, as seen in most existing data augmentation methods. This innovative method not only solves the fallaciousness due to conventional data augmentation, but also improves the diversity of training samples, and adds random noise to them. Experiments verify the effectiveness of the method. Particularly, the method is very useful to reduce false positives. To the best of our knowledge, this is the first time that only target features are transformed for data augmentation.

Following Research Question 2, how to automatically generate training for deep learning is investigated. **Research Question 3** is addressed by utilizing symbolization knowledge to automatically create training data for road extraction from the Siegfried map with U-Net. The essential idea is to first reconstruct symbols from the map, and then apply them to symbolize contemporary geo-spatial data, such as roads, railways, buildings, forests. In this way, a new map named “imitation map” is generated.

Annotations of the imitation map are inherently available (e.g., road ground-truth). Thus, it can directly be used for training. As roads are on the black layer of the map, symbol reconstruction is narrowed down to the black layer, i.e., all black symbols are reconstructed. Four training scenarios are designed and implemented, i.e., training with the original Siegfried map, training with the black layer of the Siegfried map, training with the imitation map only and training with mixed Siegfried map and imitation map. In each training scenario, the numbers of training samples are varied to investigate the impact of the amount of training samples. To the best of our knowledge, this is the first time that symbol reconstruction is used for creating training data. The method is intuitive, effective and efficient. The comparison between results of the four training scenarios reveals that this method is particularly useful in two cases. First, models trained with only the imitation map can already yield satisfactory results. Second, mixing synthetic and real training data results in models that even surpass those trained with only real data for some metrics.

To tackle **Research Question 4**, a comprehensive framework for producing raster road segmentations, vectorized road centerlines and road classifications is proposed and implemented based on deep learning and symbol painting. Road segmentations are generated with U-Net. Road skeletons are produced with morphological operations. The essence of vectorization is to find a set of connected segments that can closely approximate the road skeleton curve. For classification, road classes are distinguished with road symbols on the map, i.e., each road class corresponds to a road symbol. Thus, the key of classifying each vectorized segment is to find its corresponding symbol representation on the map. Therefore, to implement vectorization and classification, the Douglas-Peucker algorithm is first employed to find simplification/generalization points on each road skeleton curve. A sequence of connected segments approximating each road skeleton curve is generated by linking these points. For classification, a painting function is defined to paint the symbol of a road class. In this use case, five painting functions are defined. For a segment, each of the five painting functions are applied to symbolize it, and a loss is computed for the painted/symbolized segment against the target, i.e., its correspondence on the black layer of the map. The symbol/painting function that achieves a minimum loss is used to label the class of the segment. In this way, vectorization and classification are achieved in one step. To the best of our knowledge, this is the first time that a painting method is utilized to classify roads.

Although the methodologies are validated on road extraction from the Siegfried map in this thesis, they can also be applied to other feature extraction tasks (e.g., vegetation, railway) and to other historical map series (e.g., Swiss Old National map).

## 7.2. Outlook for Future Research

### 7.2.1. Method-oriented directions

First, U-Net models perform well in semantic segmentation and object detection. However, roads are long slender lines or areas in historical maps. Thus, segmented roads are subjected to discontinuity, and no topological information of the road network is learnt and preserved. Therefore, more advanced deep learning architectures should be investigated and used. Inspirations are drawn from the studies on overhead imagery. For example, atrous CNN, also known as dilated CNN, introduces dilation convolutions, which allow CNN to capture multi-scale contextual information efficiently. This architecture has been used for road extraction from overhead imagery, and the connectivity of segmented roads is improved (He *et al.*, 2019). Tao *et al.* (2019) design and integrate a spatial information inference structure (SIIS) into a convolutional recurrent neural network (RNN), which is able to capture multidirectional spatial contextual information and transmit it between pixels. The continuity of the segmented roads is considerably improved, as road-specific contextual information is captured and transmitted. Li *et al.* (2019) develop an end-to-end deep learning architecture, named PolyMapper. It

directly produces topological maps of roads and building footprints from overhead imagery. These architectures need to be adapted and applied to road extraction from historical maps to preserve the continuity and topology of roads. Additionally, the self-attention mechanism in the Transformer architecture enables it to capture long-range dependencies and to model complex relationships in input data (Vaswani *et al.*, 2017). Since roads can extend over very large areas, it should be investigated how Transformer can be used to extract roads from historical maps.

Second, this thesis proposes to use data augmentation and symbol reconstruction to solve the issue of scarcity of training data. Another way to tackle this is transfer learning, which transfers knowledge or learnt representations from a foundation model to a specific downstream task (Zoph *et al.*, 2020; Huo *et al.*, 2023). A foundation model refers to a pre-trained, general-purpose model that serves as the starting point for various specific tasks. These models are typically trained on massive datasets and are designed to capture a broad range of information and patterns from the data (Wu *et al.*, 2023). In transfer learning, a foundation model can be fine-tuned to perform well on a more specific task. For example, a foundation model trained on road data of overhead imagery or other historical maps series can be transferred or fine-tuned to another historical map series. Alternatively, more generally, a foundation model can be trained on massive overhead imagery or map data. Then, it can be fine-tuned for various down-stream segmentation tasks from historical maps. One example of a foundation model for earth observation is provided by NASA and IBM<sup>41</sup>. Furthermore, a comparison between transfer learning, data augmentation and symbol reconstruction needs to be made.

Third, in Chapter 5, symbol reconstruction is employed to construct symbols on the Siegfried map. To enhance this approach, it can be worth exploring the idea of developing a neural network that is able to analyze the legend on a map, and then to automatically reconstruct the symbols that can be used for the imitation map. In this way, the manual construction process can be avoided.

Due to quality issues in historical maps (as stated in Chapter 2), roads of the same class may have inconsistent widths. Widths and colours of road edges of the same class may slightly vary. Figure 7.1 shows an example, where the width of the road pointed by a red arrow is smaller than other roads, and the road contour pointed by a blue arrow is thicker than others, although they represent the same road class. Varying parameter values in the painting functions in Chapter 6 can solve the issues. However, designing and implementing a neural network that is able to learn these inconsistent representations of road symbols and to tolerate the inconsistencies will be more elegant.

---

<sup>41</sup> <https://www.earthdata.nasa.gov/news/impact-ibm-hls-foundation-model>, (accessed on 12.10.2023)

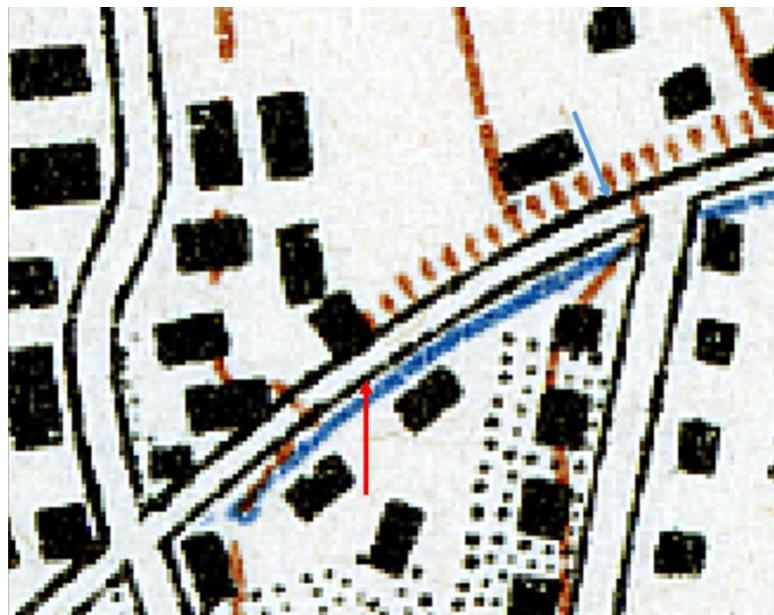


Figure 7.1. An example showing inconsistencies in road widths and contour width. Geodata © Swissstopo

Lastly, a feature extraction task usually requires a dedicated method. For example, methods for road extraction and vegetation extraction are different. Neural rendering is used for synthesizing images or videos based on neural networks. These neural networks are trained on a large number of images to learn patterns, textures, and lighting conditions, allowing them to reproduce these elements in synthetic renderings (Tewari *et al.*, 2022). Specially, neural rendering can be used for reproducing 2D historical map images. It can be trained for rendering lines and polygons, and potentially for drawing symbols (Kirillov *et al.*, 2020). In this process, it provides parameters such as feature classes (roads, buildings), locations of the start and end points of a road/railway segment, locations of the corner points of a building, etc. In this way, all features on historical maps can be vectorized and classified under one frame.

### 7.2.2. Application-oriented directions

The extracted road data and the developed methods can also be applied and expanded to the research in some other fields. First, long-term historical road data can be used for spatio-temporal analysis to unveil the evolution pattern of road network and its elementary mechanism. For example, network measures (e.g., centrality, connectivity) are computed and compared across time to analyze the evolution of the road network in Milan from 1833 to 2007 (Strano *et al.*, 2012). The results indicate that the growth of the road network is governed by two elementary processes: (1) “densification”, which involves the augmentation of local road density around established urban centers, and (2) “exploration”, where the emergence of new roads drives the spatial progression of urbanization. Likewise, road evolutions can present different patterns at global and local scales. In this case, global scale refers to Switzerland as a whole, while local scale refers to its local regions. The common changes that took place across the whole country, such as a general exploration and densification process, present homogeneous evolution patterns, whereas specific local changes, such as the merging of expanded residential centers, present heterogeneous patterns (Strano *et al.*, 2012). Furthermore, the patterns of evolutions can be compared with those of other countries (e.g., Germany). The road network evolution results can provide valuable

insights into the dynamics of human settlements and urban growth, and help shape more sustainable and resilient communities.

Second, the extracted historical roads allow urban planners to analyze historical transportation infrastructures and quantify their influences on the demographics and economy. As an important component of national, regional, and local transportation infrastructure, road networks can reflect potential walkability, accessibility, and sustainability of cities (Uhl *et al.*, 2022). Improved accessibility and sustainability of cities often leads to changes in population demographics and economic activity. For instance, Fuhrer (2020) constructed the historical transportation networks from 1720 to 2010 in Switzerland and neighboring regions to calculate historical accessibility metrics for different time periods. Li *et al.* (2021) explored the historical transportation accessibility of the Chinese Sui-Tang Period (Year 609-813) and its socioeconomic influence in the long term. It is found that disruptive occurrences such as natural disasters and wars have historically been the main drivers of population dynamics in ancient times, and accessibility can magnify this phenomenon, either hastening population concentration or precipitating population decline. This indicates that the long-term historical road data also opens an avenue for understanding socioeconomic processes in ancient times through the lens of transportation infrastructures.

Lastly, the extracted historical road data can also be used to investigate and explain urban and regional development by combining them with other related data sources. Urban and regional development is the result of interaction among various factors including road networks. For example, Lai *et al.* (2023) examined the relationship between population concentration and transport infrastructure development by considering the railway network density, and the road network density of 2'833 municipalities in Switzerland during the years 1910-2000 with a graph convolutional neural network (GCN) approach. It is found that the effects of geographical and economic constraints, as well as urban sprawl, on regional development show considerable variation among municipalities, with constraints being particularly significant during the early 20<sup>th</sup> century. Moreover, historical data on rail transit and arterial road expansions are gathered from old maps and some other sources to explore the urban transportation infrastructure development in São Paulo from 1947 until 1997 (Costa *et al.*, 2021). It is reported that every kilometer of newly constructed avenues and arterial roads resulted in a local urbanization rate increase ranging from 5% to 9%. Additionally, the historical road data can be used together with other data (e.g., human settlements) to investigate how roads facilitate interactions between settlements, which provides useful insight to the development of settlements. (Räth *et al.*, 2023). In summary, road network usually represents a key driver of socio-economic urban processes. Long-term historical road data are capable of effectively revealing urban and regional development.

## Reference

- Affek, A., 2013. Georeferencing of historical maps using GIS, as exemplified by the Austrian Military Surveys of Galicia. *Geographia Polonica*, 86(4), pp.375-390. <https://doi.org/10.7163/gpol.2013.30>
- Albu, E., 2014. *The medieval Peutinger map*. Cambridge University Press.
- Arlitsch, K., 2002. Digitizing Sanborn Fire Insurance Maps for a full color, Publicly Accessible Collection. *D-Lib magazine*, 8(7/8). <https://doi.org/10.1045/july2002-arlitsch>
- Atzori, L., Iera, A. and Morabito, G., 2010. The internet of things: A survey. *Computer networks*, 54(15), pp.2787-2805. <https://doi.org/10.1016/j.comnet.2010.05.010>
- Banister, D. and Berechman, J., 2003. *Transport investment and economic development*. Routledge.

Batty, M., Morphet, R., Masucci, P. and Stanilov, K., 2014. Entropy, complexity, and spatial information. *Journal of geographical systems*, 16, pp.363-385. <https://doi.org/10.1007/s10109-014-0202-2>

Baty, J.W., Maitland, C.L., Minter, W., Hubbe, M.A. and Jordan-Mowery, S.K., 2010. Deacidification for the conservation and preservation of paper-based works: a review. *BioResources*, 5(3). <https://doi.org/10.15376/biores.5.3.1955-2023>

Bawa, V.S. and Kumar, V., 2019. Linearized sigmoidal activation: A novel activation function with tractable non-linear characteristics to boost representation capability. *Expert Systems with Applications*, 120, pp.346-356. <https://doi.org/10.1016/j.eswa.2018.11.042>

Beineke, D., 2001. *Verfahren zur Genauigkeitsanalyse für Altkarten*. Univ. Bundeswehr München.

Biszak, E., Biszak, S., Timár, G., Nagy, D. and Molnár, G., 2017, April. Historical topographic and cadastral maps of Europe in spotlight-Evolution of the MAPIRE map portal. In *Proceedings of the 12th ICA Conference Digital Approaches to Cartographic Heritage, Venice, Italy* (pp. 26-28).

Breunig, M. and Zlatanova, S., 2011. 3D geo-database research: Retrospective and future directions. *Computers & Geosciences*, 37(7), pp.791-803. <https://doi.org/10.1016/j.cageo.2010.04.016>

Briegle, K. and Hanebeck, U.D., 2001, March. Template matching using fast normalized cross correlation. In *Optical Pattern Recognition XII* (Vol. 4387, pp. 95-102). SPIE. <https://doi.org/10.1117/12.421129>

Brown, P.G., Laycock, S.D. and Day, A.M., 2012. Vectorising Building footprints from historic maps.

Brunelli, R., 2009. Template matching techniques in computer vision: theory and practice. John Wiley & Sons.

Brüngger, A., 2011. Gletscher im Märtental: Von der Erstausgabe der Landeskarte zum dreidimensionalen Blockbild (Master thesis, ETH Zurich).

Budig, B., 2018. Extracting spatial information from historical maps: algorithms and interaction. BoD-Books on Demand.

Burrough, P.A., 1996. Natural objects with indeterminate boundaries. Burrough, PA und AU Frank (Hrsg.): *Geographic objects with indeterminate boundaries*, pp.3-28.

<https://doi.org/10.1201/9781003062660-2>

Camps-Valls, G., Tuia, D., Gómez-Chova, L., Jiménez, S. and Malo, J., 2011. Remote sensing image processing. <https://doi.org/10.1007/978-3-031-02247-0>

Canny, J., 1986. A computational approach to edge detection. *IEEE Transactions on pattern analysis and machine intelligence*, (6), pp.679-698. <https://doi.org/10.1109/tpami.1986.4767851>

Chaudhuri, G. and Clarke, K.C., 2015. On the spatiotemporal dynamics of the coupling between land use and road networks: does political history matter? *Environment and Planning B: Planning and Design*, 42(1), pp.133-156. <https://doi.org/10.1068/b39089>

Chazalon, J., Carlinet, E., Chen, Y., Perret, J., Duménieu, B., Mallet, C., Géraud, T., Nguyen, V., Nguyen, N., Baloun, J. and Lenc, L., 2021, September. ICDAR 2021 competition on historical map segmentation. In *International Conference on Document Analysis and Recognition* (pp. 693-707). Cham: Springer International Publishing. [https://doi.org/10.1007/978-3-030-86337-1\\_46](https://doi.org/10.1007/978-3-030-86337-1_46)

Chellapilla, K., Puri, S. and Simard, P., 2006, October. High performance convolutional neural networks for document processing. In *Tenth international workshop on frontiers in handwriting recognition*. Suvisoft.

Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K. and Yuille, A.L., 2017. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, 40(4), pp.834-848.

<https://doi.org/10.1109/tpami.2017.2699184>

Chen, J., Lei, Q., Miao, Y. and Peng, Q., 2015. Vectorization of line drawing image based on junction analysis. *Sci. China Inf. Sci.*, 58(7), pp.1-14. <https://doi.org/10.1007/s11432-014-5246-x>

Chen, X. and Tung, Y.K., 2003. Investigation of polynomial normal transform. *Structural Safety*, 25(4), pp.423-445. [https://doi.org/10.1016/s0167-4730\(03\)00019-5](https://doi.org/10.1016/s0167-4730(03)00019-5)

Chen, Y., Wang, R. and Qian, J., 2006. Extracting contour lines from common-conditioned topographic maps. *IEEE Transactions on Geoscience and Remote Sensing*, 44(4), pp.1048-1057.

<https://doi.org/10.1109/tgrs.2005.861478>

Cheng, H.D., Jiang, X.H., Sun, Y. and Wang, J., 2001. Color image segmentation: advances and prospects. *Pattern recognition*, 34(12), pp.2259-2281. [https://doi.org/10.1016/s0031-3203\(00\)00149-7](https://doi.org/10.1016/s0031-3203(00)00149-7)

Cheng, J., Wu, Y., AbdAlmageed, W. and Natarajan, P., 2019. QATM: Quality-aware template matching for deep learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 11553-11562). <https://doi.org/10.1109/cvpr.2019.01182>

Chiang, Y.Y., 2010. *Harvesting geographic features from heterogeneous raster maps*. University of Southern California.

Chiang, Y.Y., 2015, November. Querying historical maps as a unified, structured, and linked spatiotemporal source: vision paper. In *Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems* (pp. 1-4). <https://doi.org/10.1145/2820783.2820887>

Chiang, Y.Y., 2017. Unlocking textual content from historical maps-potentials and applications, trends, and outlooks. In *Recent Trends in Image Processing and Pattern Recognition: First International Conference, RTIP2R 2016, Bidar, India, December 16–17, 2016, Revised Selected Papers 1* (pp. 111-124). Springer Singapore. [https://doi.org/10.1007/978-981-10-4859-3\\_11](https://doi.org/10.1007/978-981-10-4859-3_11)

Chiang, Y.Y., Duan, W., Leyk, S., Uhl, J.H. and Knoblock, C.A., 2020a. *Using historical maps in scientific studies: Applications, challenges, and best practices* (pp. 10-25). Cham: Springer.

Chiang, Y.Y., Duan, W., Leyk, S., Uhl, J.H., Knoblock, C.A., Chiang, Y.Y., Duan, W., Leyk, S., Uhl, J.H. and Knoblock, C.A., 2020b. Creating structured, linked geographic data from historical maps: challenges and trends. *Using Historical Maps in Scientific Studies: Applications, Challenges, and Best Practices*, pp.37-63. [https://doi.org/10.1007/978-3-319-66908-3\\_3](https://doi.org/10.1007/978-3-319-66908-3_3)

Chiang, Y.Y. and Knoblock, C., 2013, August. Strabo: a complete system for label recognition in maps. In *26th International Cartographic Conference, Dresden, Germany* (pp. 25-30).

Chiang, Y.Y. and Knoblock, C.A., 2015. Recognizing text in raster maps. *GeoInformatica*, 19, pp.1-27. <https://doi.org/10.1007/s10707-014-0203-9>

Christophe, S., Mermet, S., Laurent, M. and Touya, G., 2022. Neural map style transfer exploration with GANs. *International Journal of Cartography*, 8(1), pp.18-36.

<https://doi.org/10.1080/23729333.2022.2031554>

Comer, M.L. and Delp III, E.J., 1999. Morphological operations for color image processing. *Journal of electronic imaging*, 8(3), pp.279-289. <https://doi.org/10.1117/1.482677>

Congalton, R.G., 1997. Exploring and evaluating the consequences of vector-to-raster and raster-to-vector conversion. *Photogrammetric Engineering and Remote Sensing*, 63(4), pp.425-434.

Costa, A.B., Zegras, P.C. and Zheng, S., 2021. Roads, transit, and the denseness of São Paulo's urban development. *MIT Center for Real Estate Research Paper*, (21/02).  
<https://doi.org/10.2139/ssrn.3784751>

Crassard, R., Abu-Azizeh, W., Barge, O., Brochier, J.É., Preusser, F., Seba, H., Kiouche, A.E., Régagnon, E., Sánchez Priego, J.A., Almalki, T. and Tarawneh, M., 2023. The oldest plans to scale of humanmade mega-structures. *Plos one*, 18(5), p.e0277927.  
<https://doi.org/10.1371/journal.pone.0277927>

Cowell, M., Leatherman, S.P. and Buckley, M.K., 1991. Historical shoreline change: error analysis and mapping accuracy. *Journal of coastal research*, pp.839-852.

Daniil, M., Tsionkas, V., Papadopoulos, K. and Livieratos, E., 2003. Scanning options and choices in digitizing historic maps. *New perspectives to save cultural heritage*.

Ester, M., Kriegel, H.P., Sander, J. and Xu, X., 1996, August. A density-based algorithm for discovering clusters in large spatial databases with noise. In *kdd* (Vol. 96, No. 34, pp. 226-231).

ETH Library, 2023. Geodata, maps, aerial photos. Available in <https://library.ethz.ch/en/locations-and-media/media-types/geodata-maps-aerial-photos.html> (Last access on 10.April 2023).

Farella, E.M., Özdemir, E. and Remondino, F., 2021. 4D building reconstruction with machine learning and historical maps. *Applied Sciences*, 11(4), p.1445. <https://doi.org/10.3390/app11041445>

Fischler, M.A. and Bolles, R.C., 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6), pp.381-395. <https://doi.org/10.1145/358669.358692>

Fleet, C., Kowal, K.C. and Pridal, P., 2012. Georeferencer: Crowdsourced georeferencing for map library collections. *D-Lib magazine*, 18(11/12), p.52. <https://doi.org/10.1045/november2012-fleet>

Forstner, G., 1998. Zwei Konstruktionsmethoden von Verzerrungsgittern zur Untersuchung alter Karten. *Cartographica Helvetica*, 18, pp.33-40.

Führer, R. and Axhausen, K.W., 2018. Reconstruction of Western Europe's pre-motorway road network back to 1500: Data sources, historical information processing, and first application. In *2018 TRB Annual Meeting Online* (pp. 18-05184). Transportation Research Board.

Führer, R., 2020. Modelling historical accessibility and its effects in space (Doctoral dissertation, ETH Zurich).

Fukunaga, K. and Hostetler, L., 1975. The estimation of the gradient of a density function, with applications in pattern recognition. *IEEE Transactions on information theory*, 21(1), pp.32-40. <https://doi.org/10.1109/tit.1975.1055330>

Ghircoias, T. and Brad, R., 2011. Contour lines extraction and reconstruction from topographic maps. *Ubiquitous Computing and Communication Journal*, 6(2), pp.681-691.

Gkonos, C., Iosifescu Enescu, I. and Hurni, L., 2019. Spinning the wheel of design: evaluating geoportal graphical user interface adaptations in terms of human-centred design. *International Journal of Cartography*, 5(1), pp.23-43. <https://doi.org/10.1080/23729333.2018.1468726>

Gobbi, S., Ciolli, M., La Porta, N., Rocchini, D., Tattoni, C. and Zatelli, P., 2019. New tools for the classification and filtering of historical maps. *ISPRS International Journal of Geo-Information*, 8(10), p.455. <https://doi.org/10.3390/ijgi8100455>

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. and Bengio, Y., 2020. Generative adversarial networks. *Communications of the ACM*, 63(11), pp.139-144. <https://doi.org/10.1145/3422622>

- Griffin, E. and Lipkin, R., 2018. 10 How Metadata informs purpose: A case study of NYPL's open source GIS-driven mapwarper tool. *Organization, Representation and Description through the Digital Age: Information in Libraries, Archives and Museums*, p.160. <https://doi.org/10.1515/9783110337419-011>
- Grosjean, G., 2013. *Geschichte der Kartographie* (Vol. 8). Geographica Bernensia.
- Gugerli, D. and Speich, D., 2002. *Topografien der Nation: Politik, kartografische Ordnung und Landschaft im 19. Jahrhundert*. Chronos-Verlag. <https://doi.org/10.1017/s0395264900003073>
- Haklay, M., 2010. How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets. *Environment and planning B: Planning and design*, 37(4), pp.682-703. <https://doi.org/10.1068/b35097>
- Hamad, K. and Mehmet, K., 2016. A detailed analysis of optical character recognition technology. *International Journal of Applied Mathematics Electronics and Computers*, (Special Issue-1), pp.244-249. <https://doi.org/10.18100/ijamec.270374>
- Han, Y., 2021. Reliable template matching for image detection in vision sensor systems. *Sensors*, 21(24), p.8176. <https://doi.org/10.3390/s21248176>
- Haque, A.U., 2018. Muhammad al-Idrisi Father of Modern Geography and Maps!. *International Journal of Pathology*, pp.98-100.
- Harley, J.B., 1988. Silences and secrecy: the hidden agenda of cartography in early modern Europe. *Imago mundi*, 40(1), pp.57-76. <https://doi.org/10.1080/03085698808592639>
- Hashemi, N.S., Aghdam, R.B., Ghiasi, A.S.B. and Fatemi, P., 2016. Template matching advances and applications in image analysis. *arXiv preprint arXiv:1610.07231*.
- Herrault, P.A., Sheeren, D., Fauvel, M. and Paegelow, M., 2013. Automatic extraction of forests from historical maps based on unsupervised classification in the CIELab color space. *Geographic information science at the heart of Europe*, pp.95-112. [https://doi.org/10.1007/978-3-319-00615-4\\_6](https://doi.org/10.1007/978-3-319-00615-4_6)
- Höhn, W., Schmidt, H.G. and Schöneberg, H., 2013, July. Semiautomatic recognition and georeferencing of places in early maps. In *Proceedings of the 13th ACM/IEEE-CS joint conference on Digital libraries* (pp. 335-338). <https://doi.org/10.1145/2467696.2467734>
- Höhn, W. and Schommer, C., 2017. Georeferencing of place markers in digitized early maps by using similar maps as data source. In *Digital Humanities 2017: Conference Abstracts*.
- Hughes, T.J., 2012. *The finite element method: linear static and dynamic finite element analysis*. Courier Corporation.
- Huo, Z., Sim, K.C., Li, B., Hwang, D., Sainath, T.N. and Strohman, T., 2023, June. Resource-efficient transfer learning from speech foundation model using hierarchical feature fusion. In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 1-5). IEEE. <https://doi.org/10.1109/icassp49357.2023.10096799>
- Hurni, L. and Brandenberger, C., 2001. Vektordaten und Optimierung der Nachführung der Landeskarte 1: 25,000. *Vorstudie*. Eidgenössische Technische Hochschule, Zürich.
- Imhof, E., 1927. Unsere Landkarten und ihre weitere Entwicklung. In: *Schweizerischer Geometerverein, ed. Zeitschrift für Vermessungswesen und Kulturtechnik*. Buchdruckerei Winterthur. S. 81–178. Available from: <https://www.e-periodica.ch/digbib/view?pid=geo-003:1927:25::440#87>
- Jenny, B. and Hurni, L., 2011. Studying cartographic heritage: Analysis and visualization of geometric distortions. *Computers & Graphics*, 35(2), pp.402-411. <https://doi.org/10.1016/j.cag.2011.01.005>

- Jenny, B., Weber, A. and Hurni, L., 2007. Visualizing the planimetric accuracy of historical maps with MapAnalyst. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 42(1), pp.89-94. <https://doi.org/10.3138/carto-v42-1-089>
- Jiao, C., Heitzler, M. and Hurni, L., 2024. A novel framework for road vectorization and classification from historical maps based on deep learning and symbol painting. *Computers, Environment and Urban Systems*, 108, p.102060. <https://doi.org/10.1016/j.compenvurbsys.2023.102060>
- Khotanzad, A. and Zink, E., 1996, April. Colour paper map segmentation using eigenvector line-fitting. In *Proceeding of Southwest Symposium on Image Analysis and Interpretation* (pp. 190-194). IEEE. <https://doi.org/10.1109/iai.1996.493751>
- Khotanzad, A. and Zink, E., 2003. Contour line and geographic feature extraction from USGS colour topographical paper maps. *IEEE transactions on pattern analysis and machine intelligence*, 25(1), pp.18-31. <https://doi.org/10.1109/tpami.2003.1159943>
- Kirillov, A., Wu, Y., He, K. and Girshick, R., 2020. Pointrend: Image segmentation as rendering. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 9799-9808). <https://doi.org/10.1109/cvpr42600.2020.00982>
- Kodama, M., 2013. Mapping downtown Honolulu's urban past: exploring the potential of geographic information systems (GIS) and Sanborn Fire Insurance Maps for historic preservation.
- Kowal, K.C. and P [rbreve] idal, P., 2012. Online georeferencing for libraries: the British Library implementation of georeferencer for spatial metadata enhancement and public engagement. *Journal of Map & Geography Libraries*, 8(3), pp.276-289. <https://doi.org/10.1080/15420353.2012.700914>
- Lai, Y., Sun, W., Schmöcker, J.D., Fukuda, K. and Axhausen, K.W., 2023. Explaining a century of Swiss regional development by deep learning and SHAP values. *Environment and Planning B: Urban Analytics and City Science*, 50(8), pp.2238-2253. <https://doi.org/10.1177/23998083221116895>
- Landini, G., Galton, A., Randell, D. and Fouad, S., 2019. Novel applications of discrete mereotopology to mathematical morphology. *Signal Processing: Image Communication*, 76, pp.109-117. <https://doi.org/10.1016/j.image.2019.04.018>
- Laumer, D., Gümgümü, H., Heitzler, M. and Hurni, L., 2020, March. A semi-automatic label digitization workflow for the Siegfried Map. In *Automatic Vectorisation of Historical Maps, Proceedings of the International Workshop on Automatic Vectorisation of Historical Maps, Budapest, Hungary* (Vol. 13, pp. 55-62). <https://doi.org/10.21862/avhm2020.07>
- LeCun, Y., Bengio, Y. and Hinton, G., 2015. Deep learning. *Nature*, 521(7553), pp.436-444. <https://doi.org/10.1038/nature14539>
- LeCun, Y., Bottou, L., Bengio, Y. and Haffner, P., 1998. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), pp.2278-2324. <https://doi.org/10.1109/5.726791>
- Lederle, C., 2017. Learning beyond the original purpose with Sanborn Fire Insurance Maps. *Social Education*, 81(6), pp.351-355.
- Levin, N., Kark, R. and Galilee, E., 2010. Maps and the settlement of southern Palestine, 1799–1948: an historical/GIS analysis. *Journal of Historical Geography*, 36(1), pp.1-18. <https://doi.org/10.1016/j.jhg.2009.04.001>
- Levin, N. and Kark, S., 2023. From historical maps to remote sensing: reconstructing land use changes on Norfolk Island over the past 250 years. *The Cartographic Journal*, pp.1-22. <https://doi.org/10.1080/00087041.2022.2150367>
- Lewis, J.P., 1995, May. Fast template matching. In *Vision interface* (Vol. 95, No. 120123, pp. 15-19).

- Leyk, S. and Boesch, R., 2009. Extracting composite cartographic area features in low-quality maps. *Cartography and Geographic Information Science*, 36(1), pp.71-79. <https://doi.org/10.1559/152304009787340115>
- Li, W., Schmöcker, J.D., Qureshi, A.G. and Zhao, L., 2021. Historical transportation accessibility of Chinese Sui-Tang period and its socioeconomics influence. In *Proceedings of the Eastern Asia Society for Transportation Studies* (Vol. 13).
- Li, Y. and Briggs, R., 2012. An automated system for image-to-vector georeferencing. *Cartography and Geographic Information Science*, 39(4), pp.199-217. <https://doi.org/10.1559/152304063941199>
- Li, Z., Kim, J., Chiang, Y.Y. and Chen, M., 2022. SpaBERT: A pretrained language model from geographic data for geo-entity representation. *arXiv preprint arXiv:2210.12213*.
- Li, Z., Wegner, J.D. and Lucchi, A., 2019. Topological map extraction from overhead images. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 1715-1724). <https://doi.org/10.1109/iccv.2019.00180>
- Liu, D., Toman, E., Fuller, Z., Chen, G., Londo, A., Zhang, X. and Zhao, K., 2018. Integration of historical map and aerial imagery to characterize long-term land-use change and landscape dynamics: An object-based analysis via Random Forests. *Ecological indicators*, 95, pp.595-605. <https://doi.org/10.1016/j.ecolind.2018.08.004>
- Liu, L., Wu, F.X., Wang, Y.P. and Wang, J., 2020. Multi-receptive-field CNN for semantic segmentation of medical images. *IEEE Journal of Biomedical and Health Informatics*, 24(11), pp.3215-3225. <https://doi.org/10.1109/jbhi.2020.3016306>
- Liu, C., Wu, J., Kohli, P. and Furukawa, Y., 2017. Raster-to-vector: Revisiting floorplan transformation. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 2195-2203). <https://doi.org/10.1109/iccv.2017.241>
- Livieratos, E., 1998. Cartographic heritage enhancement using new technologies—‘CartoTech’. *Concluding Report, 1999*.
- Livieratos, E., 2006. On the study of the geometric properties of historical cartographic representations. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 41(2), pp.165-176. <https://doi.org/10.3138/rm86-3872-8942-61p4>
- Lloyd, S., 1982. Least squares quantization in PCM. *IEEE transactions on information theory*, 28(2), pp.129-137. <https://doi.org/10.1109/tit.1982.1056489>
- Loran, C., Haegi, S. and Ginzler, C., 2018. Comparing historical and contemporary maps—a methodological framework for a cartographic map comparison applied to Swiss maps. *International Journal of Geographical Information Science*, 32(11), pp.2123-2139. <https://doi.org/10.1080/13658816.2018.1482553>
- Luft, J. and Schiewe, J., 2021. Automatic content-based georeferencing of historical topographic maps. *Transactions in GIS*, 25(6), pp.2888-2906. <https://doi.org/10.1111/tgis.12794>
- Masucci, A.P., Arcaute, E., Wang, J., Hatna, E., Stanilov, K. and Batty, M., 2015. Logistic growth and ergodic properties of urban forms. *arXiv preprint arXiv:1504.07380*.
- Masucci, A.P., Stanilov, K. and Batty, M., 2013. Limited urban growth: London’s street network dynamics since the 18<sup>th</sup> century. *PLoS one*, 8(8), p.e69469. <https://doi.org/10.1371/journal.pone.0069469>

- Medina-Carnicer, R., Carmona-Poyato, A., Muñoz-Salinas, R. and Madrid-Cuevas, F.J., 2009. Determining hysteresis thresholds for edge detection by combining the advantages and disadvantages of thresholding methods. *IEEE transactions on image processing*, 19(1), pp.165-173.  
<https://doi.org/10.1109/tip.2009.2032942>
- Mello, C.A., Costa, D.C. and Dos Santos, T.J., 2012, October. Automatic image segmentation of old topographic maps and floor plans. In *2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC)* (pp. 132-137). IEEE. <https://doi.org/10.1109/icsmc.2012.6377689>
- Mitchell, T.M., 1997. Machine Learning-McGraw-Hill Science. Engineering.
- Muller, J.C. ed., 1991. *Advances in cartography*. International Cartographic Association by Elsevier Applied Science.
- Mueller, L., 2004. Sanborn Fire Insurance Maps: history, use, availability. *The Primary Source*, 26(2), p.2. <https://doi.org/10.18785/ps.2602.02>
- Mueller, L., 2004. Sanborn Fire Insurance Maps of Mississippi: A list. *The Primary Source*, 26(2), p.3. <https://doi.org/10.18785/ps.2602.03>
- Murcio, R., Masucci, A.P., Arcaute, E. and Batty, M., 2015. Multifractal to monofractal evolution of the London street network. *Physical Review E*, 92(6), p.062130.  
<https://doi.org/10.1103/physreve.92.062130>
- Nair, V. and Hinton, G.E., 2010. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27<sup>th</sup> international conference on machine learning (ICML-10)* (pp. 807-814).
- Niu, Z., Zhong, G. and Yu, H., 2021. A review on the attention mechanism of deep learning. *Neurocomputing*, 452, pp.48-62. <https://doi.org/10.1016/j.neucom.2021.03.091>
- Novak, C.L. and Shafer, S.A., 1992, June. Anatomy of a colour histogram. In *CVPR* (Vol. 92, pp. 599-605). <https://doi.org/10.1109/cvpr.1992.223129>
- Panagos, P., Jones, A., Bosco, C. and Kumar, P.S., 2011. European digital archive on soil maps (EuDASM): preserving important soil data for public free access. *International Journal of Digital Earth*, 4(5), pp.434-443. <https://doi.org/10.1080/17538947.2011.596580>
- Pérez Nava, F., Sánchez Berriel, I., Pérez Morera, J., Martín Dorta, N., Meier, C. and Hernández Rodríguez, J., 2023. From maps to 3D models: reconstructing the urban landscape of San Cristóbal de La Laguna in the 16<sup>th</sup> century. *Applied Sciences*, 13(7), p.4293. <https://doi.org/10.3390/app13074293>
- Perret, J., Boffet Mas, A. and Ruas, A., 2009, November. Understanding urban dynamics: the use of vector topographic databases and the creation of spatio-temporal databases. In *24<sup>th</sup> international cartography conference (icc'09)*.
- Perret, J., Gribaudi, M. and Barthelemy, M., 2015. Roads and cities of 18<sup>th</sup> century France. *Scientific data*, 2(1), pp.1-7. <https://doi.org/10.1038/sdata.2015.48>
- Plewe, B., 2002. The nature of uncertainty in historical geographic information. *Transactions in GIS*, 6(4), pp.431-456. <https://doi.org/10.1111/1467-9671.00121>
- Powers, K.E., Prather, L.A., Cook, J.A., Woolley, J., Bart Jr, H.L., Monfils, A.K. and Sierwald, P., 2014. Revolutionizing the use of natural history collections in education. *Science Education Review*, 13(2), pp.24-33.
- Raper, J., Gartner, G., Karimi, H. and Rizos, C., 2007. Applications of location-based services: a selected review. *Journal of Location Based Services*, 1(2), pp.89-111.  
<https://doi.org/10.1080/17489720701862184>

- Ray, A., Chen, Z., Gafford, B., Gifford, N., Kumar, J.J., Lamsal, A., Niehus-Staab, L., Weinman, J. and Learned-Miller, E., 2018. Historical map annotations for text detection and recognition. *Grinnell College, Grinnell, Iowa, Tech. Rep.*
- Reckziegel, M., Wrissley, D.J., Hixson, T.W. and Jänicke, S., 2021. Visual exploration of historical maps. *Digital Scholarship in the Humanities*, 36(Supplement\_2), pp.ii251-ii272.  
<https://doi.org/10.1093/lhc/fqaa059>
- Rhind, D.W., 1991. The role of the ordnance survey of Great Britain. *The Cartographic Journal*, 28(2), pp.188-199. <https://doi.org/10.1179/caj.1991.28.2.188>
- Rickenbacher, M., 2013, August. Journeys through time with the Swiss national map series. In *Proceedings of the 26<sup>th</sup> International Cartographic Conference, Dresden, Germany* (Vol. 2530).
- Rumsey, D. and Williams, M., 2002. *Historical maps in GIS* (pp. 1-17). na.
- Sahu, M. and Dash, R., 2021. A survey on deep learning: Convolution Neural Network (CNN). In *Intelligent and Cloud Computing: Proceedings of ICICCC 2019, Volume 2* (pp. 317-325). Springer Singapore. [https://doi.org/10.1007/978-981-15-6202-0\\_32](https://doi.org/10.1007/978-981-15-6202-0_32)
- San-Antonio-Gómez, C., Velilla, C. and Manzano-Agüilaro, F., 2014. Urban and landscape changes through historical maps: The Real Sitio of Aranjuez (1775–2005), a case study. *Computers, environment and urban systems*, 44, pp.47-58. <https://doi.org/10.1016/j.compenvurbsys.2013.12.001>
- Sánchez-Berriel, I., González-González, A., Pérez-Nava, F., Meier, C., Pérez-Morera, J. and Hernández-Alberto, C.R., An interactive 3D application of a house from the XVI century in San Cristóbal de Laguna as a case study for the dissemination of cultural heritage. *Proceedings of the Joint International Event 9th ARQUEOLÓGICA*, 2. <https://doi.org/10.4995/arqueologica9.2021.12061>
- Savage, J.C. and Burford, R.O., 1973. Geodetic determination of relative plate motion in central California. *Journal of Geophysical Research*, 78(5), pp.832-845.  
<https://doi.org/10.1029/jb078i005p00832>
- 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), pp.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, p.12. <https://doi.org/10.5194/agile-giss-2-12-2021>
- Sebert, L.M., 2001. Popular maps: The Ordnance Survey Popular Edition One-Inch Map of England and Wales, 1919-1926. *Geomatica*, 55(3), pp.384-387.
- Serra, J., 1983. *Image analysis and mathematical morphology*. Academic Press, Inc.
- Sherstinsky, A., 2020. Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network. *Physica D: Nonlinear Phenomena*, 404, p.132306.  
<https://doi.org/10.1016/j.physd.2019.132306>
- Shirvani, Z., Abdi, O. and Buchroithner, M.F., 2020. A new analysis approach for long-term variations of forest loss, fragmentation, and degradation resulting from road-network expansion using Landsat time-series and object-based image analysis. *Land Degradation & Development*, 31(12), pp.1462-1481. <https://doi.org/10.1002/ldr.3530>
- Shi, W., 2009. *Principles of modeling uncertainties in spatial data and spatial analyses*. CRC press.  
<https://doi.org/10.1201/9781420059281>

Skurnik, J., 2020. Authorizing geographical knowledge: John Arrowsmith, mapmaking and the mid nineteenth-century British Empire. *Journal of Historical Geography*, 69, pp.18-31.  
<https://doi.org/10.1016/j.jhg.2020.04.003>

Sobel, I., 1968. History and definition of the so-called “Sobel Operator”, more appropriately named the Sobel-Feldman Operator. 2014. URL:

[https://www.researchgate.net/publication/239398674\\_An\\_Isotropic\\_3x3\\_Image\\_Gradient\\_Operator](https://www.researchgate.net/publication/239398674_An_Isotropic_3x3_Image_Gradient_Operator)  
*First presented at the Stanford Artificial Intelligence Project (SAIL).*

Southall, H. and Pridal, P., 2012. Old maps online: enabling global access to historical mapping. *e-Perimetron*, 7(2), pp.73-81.

Strano, E., Giometto, A., Shai, S., Bertuzzo, E., Mucha, P.J. and Rinaldo, A., 2017. The scaling structure of the global road network. *Royal Society open science*, 4(10), p.170590.

<https://doi.org/10.1098/rsos.170590>

Sutton, R.S. and Barto, A.G., 2018. *Reinforcement learning: An introduction*. MIT press.

Swisstopo, 2023a. Digital Historical Maps.

<https://www.swisstopo.admin.ch/en/geodata/maps/historical.html> (Last access on 15 March 2023).

Swisstopo, 2023b. Historical maps.

<https://www.swisstopo.admin.ch/en/knowledge-facts/histcoll/historical-maps.html> (Last access on 09 April 2023).

Symington, A., Charlton, M.E. and Brunsdon, C.F., 2002. Using bidimensional regression to explore map lineage. *Computers, Environment and Urban Systems*, 26(2-3), pp.201-218.

[https://doi.org/10.1016/s0198-9715\(01\)00042-4](https://doi.org/10.1016/s0198-9715(01)00042-4)

Tan, K.S. and Isa, N.A.M., 2011. Color image segmentation using histogram thresholding-Fuzzy C-means hybrid approach. *Pattern recognition*, 44(1), pp.1-15.

<https://doi.org/10.1016/j.patcog.2010.07.013>

Thorat, C., Bhat, A., Sawant, P., Bartakke, I. and Shirasath, S., 2022. A detailed review on text extraction using optical character recognition. *ICT Analysis and Applications*, pp.719-728.

[https://doi.org/10.1007/978-981-16-5655-2\\_69](https://doi.org/10.1007/978-981-16-5655-2_69)

Tobler, W.R., 1994. Bidimensional regression. *Geographical Analysis*, 26(3), pp.187-212.

<https://doi.org/10.1111/j.1538-4632.1994.tb00320.x>

Tschopp, M., Fröhlich, P.H. and Axhausen, K.W., 2005. Accessibility and spatial development in Switzerland During the last 50 years. In *Access to destinations*. Emerald Group Publishing Limited.  
<https://doi.org/10.1108/9780080460550-017>

Tsioukas, V., Daniil, M. and Livieratos, E., 2006. Possibilities and problems in close range non-contact 1: 1 digitization of antique maps. *e-Perimetron*, 1(3), pp.230-238.

Tsorlini, A. and Hurni, L., 2015. On the georeference and reprojection of old maps in different application software: comparing methodology and results. In *10<sup>th</sup> Jubilee Commission Conference on Digital Approaches to Cartographic Heritage*.

Tsorlini, A., Iosifescu, I. and Hurni, L., 2013, August. Comparative analysis of historical maps of the Canton of Zurich–Switzerland in an interactive online platform. In *Proceedings of the 25<sup>th</sup> International Cartographic Conference, Dresden, Germany* (pp. 25-30).

Uhl, J.H., Leyk, S., Chiang, Y.Y., Duan, W. and Knoblock, C.A., 2017, July. Extracting human settlement footprint from historical topographic map series using context-based machine learning. In *8th International Conference of Pattern Recognition Systems (ICPRS 2017)* (pp. 1-6). IET.  
<https://doi.org/10.1049/cp.2017.0144>

Uhl, J.H., Leyk, S., Chiang, Y.Y., Duan, W. and Knoblock, C.A., 2018. Map archive mining: visual-analytical approaches to explore large historical map collections. *ISPRS international journal of geoinformation*, 7(4), p.148. <https://doi.org/10.3390/jigi7040148>

Uhl, J.H., Leyk, S., Li, Z., Duan, W., Shbita, B., Chiang, Y.Y. and Knoblock, C.A., 2021. Combining remote-sensing-derived data and historical maps for long-term back-casting of urban extents. *Remote sensing*, 13(18), p.3672. <https://doi.org/10.3390/rs13183672>

University Library Bern, 2023. Map collections. Available in  
[https://www.ub.unibe.ch/research/special\\_collections/map\\_collections/index\\_eng.html](https://www.ub.unibe.ch/research/special_collections/map_collections/index_eng.html) (Last access on 09 April 2023).

Utrilla, P., Mazo, C., Sopena, M.C., Martínez-Bea, M. and Domingo, R., 2009. A palaeolithic map from 13,660 calBP: engraved stone blocks from the Late Magdalenian in Abauntz Cave (Navarra, Spain). *Journal of Human Evolution*, 57(2), pp.99-111. <https://doi.org/10.1016/j.jhevol.2009.05.005>

Velikić, I.M., Rabrenović, D., Srećković-Batočanin, D. and Spahić, D., 2022. Geoheritage and mining heritage in the promotion of theme parks: An example of the National Park Đerdap (Carpathian-Balkan Thrust-And-Fold Belt, Eastern Serbia). *Geoheritage*, 14(3), p.102.  
<https://doi.org/10.1007/s12371-022-00735-3>

Veregin, H. and Hargitai, P., 1995. An evaluation matrix for geographical data quality. In *Elements of spatial data quality* (pp. 167-188). Pergamon. <https://doi.org/10.1016/b978-0-08-042432-3.50016-1>

Vervust, S., Boùuaert, M.C., De Baets, B., Van de Weghe, N. and De Maeyer, P., 2018. A study of the local geometric accuracy of Count de Ferraris's Carte de cabinet (1770s) using differential distortion analysis. *The Cartographic Journal*, 55(1), pp.16-35. <https://doi.org/10.1080/00087041.2017.1323148>

Wagner, H., 1896. *Das ratsel der kompasskarten im lichte der gesamtentwicklung der seerkarten*. Geographische Verlagshandlung Deitrich Reimer.

Weber, P., Andreassen, L.M., Boston, C.M., Lovell, H. and Kvarteg, S., 2020. An~ 1899 glacier inventory for Nordland, northern Norway, produced from historical maps. *Journal of Glaciology*, 66(256), pp.259-277. <https://doi.org/10.1017/jog.2020.3>

Weinman, J., Chen, Z., Gafford, B., Gifford, N., Lamsal, A. and Niehus-Staab, L., 2019, September. Deep neural networks for text detection and recognition in historical maps. In *2019 International Conference on Document Analysis and Recognition (ICDAR)* (pp. 902-909). IEEE.  
<https://doi.org/10.1109/icdar.2019.00149>

Weis, I., 1985. Automatische Erstellung von Verzerrungsgittern alter Karten. *Munich: Technische Universität München*.

Wiesmann, S., 2016. Gletscheroberflächenentwicklung–Aspekte der Datenakquisition und der Visualisierung (Doctoral dissertation, ETH Zurich).

Wolodtschenko, A. and Forner, T., 2007. Prehistoric and early historic maps in Europe: Conception of cd-atlas. *e-Perimetron*, 2(2), pp.114-116.

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

Wyder, S., 2006. *Grenz-, Zehnten-und Befestigungspläne der Zürcher Gebiets von Hans Conrad Gyger (1599-1674)*. Verlag Cartographica Helvetica.

Xia, X., Heitzler, M. and Hurni, L., 2022. CNN-based template matching for detecting features from historical maps. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences-ISPRS Archives*, 43(B2-2022), pp.1167-1173. <https://doi.org/10.5194/isprs-archives-xliii-b2-2022-1167-2022>

## List of Figures

Figure 2.1. Examples of Swiss historical maps that cover a section located in north-west Zurich city	9
Figure 2.2. Examples of geographical features with fuzzy borders on the Siegfried map	14
Figure 2.3. Examples of inconsistent symbol representations	15
Figure 2.4. Examples of “unknown objects” pointed by arrows	16
Figure 2.5. Examples of blurring, colour aliasing and colour inhomogeneity issues on the Siegfried map	17
Figure 2.6. An example of colour aliasing around boundaries	18
Figure 2.7. A Figure from Tsorlini <i>et al.</i> (2013), showing the distortion in Zurich on the Gyger Map (1:32,000) of 1660 compared to the same area in modern maps. Displacement vectors are represented with blue arrows	19
Figure 2.8. An example of applying Mean-Shift to a section of a Siegfried map sheet	25
Figure 2.9. An example of template matching for wetland extraction	26
Figure 2.10. A simple yet typical architecture of a deep CNN consisting of convolutional layers, max pooling layers and fully connected layers	28
Figure 3.1. (a) Yearly counts of papers on road extraction from raster maps, and (b) citation counts of these papers. The search was conducted on September 27, 2020 using Google Scholar	53
Figure 4.1. Roads on Siegfried map	71
Figure 4.2. The road training data and its buffers	72
Figure 4.3. Sampling points and the corresponding map tiles	73
Figure 4.4. (a) a map tile cropped from Siegfried map, (b) rotating the roads in (a) by 270° anti-clockwise, (c) another map tile, (d) horizontal flipping the roads in (c)	74
Figure 4.5. The U-Net architecture	75
Figure 4.6. The comparison between raster road predictions obtained without data augmentation vs. with conventional data augmentation vs. with novel data augmentation	76
Figure 4.7. Vector road centerlines extracted by the model trained with novel data augmentation	77
Figure 5.1. Different grades of roads are represented by six types of black line symbols	84
Figure 5.2. Other features represented by black symbols	85
Figure 5.3. The historical road vector data available for training	86
Figure 5.4. Workflow of the methodology	87
Figure 5.5. Reconstructed road symbols, which correspond to those in Figure 5.1	88
Figure 5.6. The spot height and labels derived from the “place name” feature class, which are represented with the same font and size as those of Siegfried map	88
Figure 5.7. (a) Random noise in buildings and road edges on Siegfried map, (b) random noise (white pixels) in the black layer (the method to obtain the black layer is elaborated in Section 5.4.3)	89

Figure 5.8. An example of the imitation map in Zurich.....	90
Figure 5.9. The road segmentation model adapted from U-Net.....	91
Figure 5.10. The sampling strategy.....	93
Figure 5.11. Extracting the black layer from Siegfried map.....	94
Figure 5.12. An example in which Result 3 outperforms the other results regarding false positives on the label.....	98
Figure 5.13. An example in which Result 3 outperforms the other results regarding false positives on the label, grid line and forest border.....	99
Figure 5.14. Training and validation losses of the four scenarios with 20,000 samples.....	100
Figure 5.15. Examples of symbols that have complex patterns.....	101
Figure 6.1. A subsection of a Siegfried map sheet covering a region in Zurich. Each red arrow points to a road class represented by a road symbol.....	110
Figure 6.2. The available road training data are road centerlines (in blue).....	110
Figure 6.3. Workflow of the methodology.....	111
Figure 6.4. (a) A section from the original Siegfried map, (b) road segmentation result, (c) road skeletons and (d) vectorized road lines.....	114
Figure 6.5. A road symbol corresponds to a road class.....	114
Figure 6.6. Two examples of skeleton curve decomposition.....	117
Figure 6.7. Examples of painting results.....	118
Figure 6.8. Examples of painting, vectorization and classification results in town/urban areas.....	121
Figure 6.9. Comparison of road vectorization (and classification) results from the painting method, as shown in the first column vs. from morphological operations, as shown in the second column.....	122
Figure 6.10. Sensitivity analysis results of road classification.....	125
Figure 7.1. An example showing inconsistencies in road widths and contour width.....	136

## List of Tables

Table 2.1. Quality issues in historical maps.....	18
Table 3.1. Categorization of road extraction methods based on raster maps.....	34
Table 4.1. Quantitative road extraction results (5,000 original samples without data augmentation)....	78
Table 4.2. Quantitative road extraction results (5,000 original samples plus 1,400 augmented samples of novel data augmentation).....	78
Table 4.3. Quantitative road extraction results (5000 original samples with additional 500 samples around road class 1, plus 1,400 augmented samples of novel data augmentation).....	78
Table 5.1. Qualitative road extraction results of the four training scenarios with 20,000 training samples.....	96
Table 5.2. Quantitative road extraction results of the four training scenarios.....	97
Table 6.1. Road segmentation results from the Siegfried map in urban areas, sub-urban areas and rural areas.....	113
Table 6.2. Parameters needed for each painting function or each road symbol.....	115
Table 6.3. The selected parameters for each painting function or each road symbol.....	119
Table 6.4. An example of comparing the loss values of painting an image patch.....	120
Table 6.5. Average metric values of the four sheets.....	123
Table 6.6. Sensitivity analysis results for different $\epsilon$ values.....	124