Extracting Land Use and Land Cover Information from the Franciscean Cadastre using Object-Based Image Analysis and Convolutional Neural Networks

Master Thesis

For obtaining the

Master's Degree

MSc

At the Faculty of Digital and Analytical Sciences

of the Paris-Lodron University Salzburg

Submitted by

Simon Zacharias Meyer

Supervisor: Dr. Dirk Tiede

Department of Geoinformatics – Z_GIS

Salzburg, March 2024

Science Pledge

I hereby declare that this work is the result of my own work. All sources used in this work have been openly presented and cited in accordance with scientific ethics.

(Rinteln, March 2024)

Abstract

Since the onset of the industrialisation in the early 19th century, land use and land cover in the province of Salzburg has changed significantly. The documentation and visualisation of this change through long-term studies is important for the evaluation of landscape changes and the core of many often interdisciplinary studies. The greatest obstacle to analysing long-term landscape change to date is the limited availability of comprehensive historical data sets. While today many geographic datasets are available, they are overwhelmingly restricted to recent timeframes and rarely older than 50 years. A proven method to extend timeframes is to utilize historic maps, but manual map vectorization is time consuming and prone to errors which has so far limited their application in the amount of vectorized historic data.

This study aimed at developing a workflow in Trimble eCognition for a semiautomatic extraction of geographic information from scanned maps using object-based image analysis and convolutional neural networks to classify and vectorize historic maps. The focus was set on the extraction of selected map symbols and land use and land cover classes from one of the most important historic map series of Austrian Empire, the Franciscean Cadastre (Franziszeischer Kataster). Results of this study should provide the foundation for potential large scale vectorization work of the cadastre and enable the quantitative analysis of the historic landscape and its transformation over 200 years.

While the goals of the study were ambitious, not all objectives could be achieved. Map symbols could not be recognized with sufficient accuracy. Therefore, they were not included in the land use and land cover classification. The surface classification achieved a 90% overall accuracy for the Salzburg region with varying accuracies per class. The overall classification accuracy dropped significantly in the second study area located in the High Tauern region, indicating a limited ability of the classification approach to correctly classify severely degraded map sections. Further work in training and parameterization is required to ensure transferability in other regions of Salzburg. Despite these limitations, the study represents a step forward in making historical geographic data available for spatial analysis and may lay the groundwork for future studies that expand the dataset geographically.

Keywords:

Historical Geographic Data, Map Classification, Franciscean Cadastre, Object-Based Image Analysis, Multiresolution Segmentation, Convolutional Neural Networks

Abbreviations

ANN – Artificial Neural Network

CNN – Convolutional Neural Network

ESP – Estimation of Scale Parameter

GIS – Geographic Information System

LULC – Land Use and Land Cover

MRS – Multiresolution Segmentation

OBIA – Object-Based Image Analysis

SVM – Support Vector Machine

RF – Random Forest

Contents

List of Figures

List of Tables

1 Introduction

1.1 Historic Map Digitization

The Anthropocene has drastically changed the Earth's landscapes. The multitude of human interventions in nature has drastically altered our environment and led to increased awareness and research on their effects on humanity. Because landscape changes often occur rather slowly and change can occur gradually over several human generations, long-term studies are crucial to further facilitate understanding about our environment (Herold, 2018).

Change can be assessed by either relying on model-based approaches, simulating the transformation based on conditions and assumptions, or by utilizing available historic documents and measurement series (Uhl et al., 2021b). Long-term land use and land cover (LULC) change studies use a variety of geoinformation sources including current hyperspectral and radar data, digital LULC maps, archival optical monochromatic airborne and multi-spectral satellite imagery, as well as older cartographic documents (Herold, 2018).

As comprehensive remote sensing mapping only really started in the early 1970s with the launch of Landsat-1, large-scale mapping products for earlier periods are lacking and data availability gets increasingly sparse as you delve deeper into the past (Uhl et al., 2021b). And if historical imagery exists, it is often severely limited in its spatial or temporal extent, being only available for selected regions. This creates problems for large-scale geographic long-term studies (Herold, 2018).

One way to reduce the issue of data availability is to use older geographic information sources, such as historic maps. In Europe, the first standardized mapping efforts began in the late 18th century (Bauer, 2017). Surveying projects documented a great variety of information that was of interest to the state such as the number of households, land ownership, agricultural yields, road networks, and military infrastructure. While these products were tailored to specific demands at the time, they today allow detailed insights into the landscape of their time and often present the only available source for LULC information, making them invaluable for long-term studies on landscape change. The problem with historic geodata is, that before using historical geographical data for quantitative analyses, it must be converted from its original paper map form into digital geodata.

The transformation of the physical map into a digital format is done by scanning the map and storing the image in a digital raster format. The raster format saves the spectral properties of the map in discrete pixel values, but as no image interpretation is conducted, no information about the contents of the image stored, which limits the usability of raster files for spatial analysis in geographic information systems (GIS), requiring the further vectorization of the image file.

The traditional way of converting historical maps into geodata consists of the visual interpretation and manual vectorization of the scanned data source by a person that has knowledge in the historic context and the available tools. The task is time and resource expensive, which has in the past often limited the spatial and temporal scope of related studies (Zatelli et al., 2022a, Ignjatić et al., 2018, Herold, 2018, Gobbi et al., 2019). It has also led to a general underutilization of historic geographic maps in geographic information science, although they are increasingly publicly available online as digitalised products, often through online databases of state archives and geoinformation offices.

Therefore, many researchers have aimed at automating this tedious task. In the past, there were multiple drivers for the development of automatic map vectorization methods, which can generally be divided into three distinct stages: With the advent of computer systems and the first artificial intelligence and recognition algorithms in the second half of the 20th century, first research was dedicated to the application on maps. Later, with the development of the first geographic information systems and geospatial analysis tools, the demand for digital spatial information rose, leading to a renewed interest for the digitization of historic data. But methods were limited by the technology of its time. Today, the growing availability of digitally available historic maps and the continuing need for historical geodata has again renewed interest in automated map vectorization methods (Herold, 2018).

1.2 Literature Review

In the last decade, many studies were published that either aimed at developing new ways to extract information from historic maps or utilized a variety of established methods to vectorize old cartographic sources for historic analysis. The following section will attempt to summarize the state of the art of both the usage of historic maps for landscape change analysis and (semi-)automatic object recognition and classification approaches used for the vectorization of historic maps.

Overall, a continuous growth in scientific literature can be observed. Chiang et al. (2020a) counted the number of published papers per year on Google Scholar containing the keywords ‘historical maps’ and ‘analysis’ between 2001 and 2018, documenting an increase from 231 to 1370 papers. The results did include papers from top-ranked scientific journals, covering several academic disciplines and domains (Chiang et al., 2020a).

Research in the field of map processing is broad and covers all parts of the map digitization process, from map scanning and georeferencing to image processing and vectorization. Chiang et al. (2014) highlighted that map processing approaches have been applied to various types of maps over the years, including cadastral or land register maps, road maps, hydrographic maps, city maps, utility maps, and topographic or other survey maps.

Most studies focused on map feature or cartographic symbol extractions on specific maps and only few research aimed at developing workflows that could be applied to all map types (Chiang et al., 2014).

An early study on the use of historic maps for LULC change analysis was conducted by Petit and Lambin (2002) who compared historical maps with remote sensing data in the Belgian Ardennes. They assessed the data integration and generalisation necessary to compare maps of different timeframes, confirming the value of historic data for long-term LULC analysis (Petit and Lambin, 2002). While they addressed cartographic challenges in comparing different map series, they did not aim at automating the vectorization process and therefore only applied their study to a rather small study area.

Shortly after, Haase et al. (2005) analysed historical landscape change in Saxony, Germany using topographic maps from the 18th to 20th century. They concluded that historical maps pose a valid source to quantitatively assess structural landscape changes and that results can be used to predict future landscape change (Haase et al., 2005).

Statuto et al. (2017) used historical maps from 1829 to 2013 in their landscape analysis of Southern Italy. They as well highlighted the importance of historical geographic data for the landscape change analysis, but as they chose a manual approach to map vectorization their study was constrained to an 18 km² area (Statuto et al., 2017). Similarly, Gimmi et al. (2011) used manually vectorized historical maps for the assessment of wetland networks in the Swiss lowlands between 1850 and 2000.

Subsequent research explored possibilities to automatically extract selected information, such as building footprints, street networks, LULC types, or text, from maps using different approaches. In addition, increased research was conducted to assess whether transferable workflows for large-scale vectorization works are possible.

Godfrey and Eveleth (2015) developed an adaptable semiautomatic map vectorization approach using unsupervised classification in ArcGIS for Desktop. A year later, Iosifescu et al. (2016) extracted building footprints and water bodies from two 19th century Swiss topographical maps based on mathematical morphology operations using GDAL, OGR library, and Image Magick.

Researchers at the University of Colorado Boulder led by S. Leyk, S. Uhl, and Y.-Y. Chiang from the University of Southern California developed multiple methods to digitize historic maps and large-scale map collections. They applied their vectorization approach to extract road networks (Uhl et al., 2022) and different land cover types (Uhl et al., 2021a) from various maps sources. They also assessed the usage of remote sensing derived data in combination with historic maps to analyse long-term urban growth (Uhl et al., 2021b).

Multiple studies centred around map digitization were also conducted by the Department of Civil, Environmental, and Mechanical Engineering at Università degli Studi di Trento and the Department of Humanities at Università degli Studi di Trento in Italy. Gobbi et al. (2019) developed new tools for the classification and filtering of historical maps. They chose a semi-automatic approach to successively filter out undesirable features, such as text, cartographic symbols, and lines. The tool was applied to several different historic maps sources and achieved high accuracies (Gobbi et al., 2019). Zatelli et al. (2019) applied the previously developed tools and workflow to a map section of the Franciscean Cadastre to assess the capabilities of OBIA for historic map classification compared to other traditional classification techniques using GRASS GIS and R. Their research concluded that 'the OBIA approach has provided very satisfactory results with the ability to automatically remove the background and symbols and characters' (Zatelli et al., 2019). In a subsequent study Zatelli et al. (2022a) developed specifically tailored OBIA workflow in GRASS GIS and R to extract thematic information from Cesare Battisti's atlas of 1915. The study emphasised the significance of historical maps in ecology and climate change research. Vectorized historical maps enabled the reconstruction of forest cover, which was deemed crucial for understanding the impact of climate change on alpine ecosystems (Zatelli et al., 2022a). Zatelli et al. (2022b) modelled the LULC change in the Trentino region based multiple historical maps from 1859 to 1936 and aerial images from 1954 to 2015 using a FOSS4G approach. High accuracies over 94% were achieved and the preliminary results enabled the creation of future forest evolution scenarios for the next 85 years (Zatelli et al., 2022b).

In Germany, Ulloa et al. (2020) created an semi-automatic map vectorization workflow to assess land cover changes in the Main river catchment, Germany using eCognition and OBIA. The approach proved to be successful in detecting and classifying the most important LULC types found in the map.

While most research used traditional classification algorithms, recent work has also tested the ability of deep neural networks for map classification and text recognition. Uhl et al. (2018) used weakly supervised CNNs to segment features in historic maps. Then, Chiang et al. (2020b) explored the capabilities deep learning models for historical map feature recognition.

Most recently, Kersapati and Grau-Bové (2023) have reviewed tested and compared the capabilities of CNNs and OBIA to extract geographic features from historic maps of the Bandu Islands, Indonesia. Results showed the superiority of CNNs to OBIA regarding statistical performance, but OBIA provided greater flexibility on multiple scales.

1.3 Current Challenges in (Semi-)Automatic Map Processing

While research on automated extraction of geospatial information from historic maps is ongoing and significant progress has been made in the last decade, the complete automatization of the raster to vector conversion process has not yet been adequately solved (Ignjatić et al., 2018). Currently, several fundamental challenges remain.

One persistent issue has been the data quality. As historic maps come in various forms and conditions, their quality is a crucial factor that must be assessed bevor any attempt of data extraction. Discolorations and paper yellowing in the original map source can lead to noise and additional challenges for recognition and classification algorithms (Herold, 2018). More quality issues can be introduced during the scanning process, where folds and holes between map sheets may cause gaps in cartographic information (Chen et al., 2021). Low resolution scans may also negatively image quality by affecting the colour grading, removing important detail, or introducing new image artifacts that could influence subsequent classification results. Much research therefore has been conducted to normalize and filter scanned historical maps in multiple pre-processing steps or to develop robust algorithms that can differentiate variations in colour and texture.

Another major issue commonly found in historic map processing is the overall complexity and structure of historic maps. Recognition algorithms can face severe difficulties when dealing with historical maps due to overlapping text, lines, and symbols, particularly in areas with high information density. Additionally, size, shape, and colour can vary (Chen et al., 2021, Zatelli et al., 2019), as symbols were often hand-drawn. The correct recognition of symbols and the extraction of text can therefore be particularly challenging, but the extent of these issue heavily depends on the type of map and its quality. Historic topographic maps for example often have low overall colour variations but many letters, lines, and text symbols. The extraction of information from these map types can therefore pose different challenges than other types of maps.

The most important factor to consider in (semi-)automated map processing is to ensure the transferability of the workflow. A robust recognition strategy is crucial to a successful map vectorization (Herold, 2018), as the process aims to keep time consuming manual adjustments to the algorithms to a minimum. The quantitative advantages of automated approaches can be best played out by extending the geographic scale of the map digitization project. The recognition and classification algorithms therefore must be able to correctly identify features in new map regions with changes to map quality and characteristics. Low generalization abilities of the classification algorithms can significantly reduce the benefits of quantitative approaches, nullifying much of the time-consuming and resource-intensive labour spent on the development of the algorithms. Achieving high

generalization abilities of the map processing approach chosen is therefore one of the main goals in automatic map vectorization projects.

Recognizing and addressing these issues is prerequisite to the generation of large-scale historical geodata collections and therefore has been a consistent goal in research in the last decades.

1.4 Objectives and Rationale of the Study

The aim of this master thesis is to develop a workflow for the automatic extraction of geographic information from the historical map series of the Franciscean Cadastre using Convolutional Neural Networks (CNNs) and object-based image analysis (OBIA).

The study has three main objectives:

1. Development of a workflow in eCognition for the extraction of map symbols, as well as the segmentation and classification of the most important LULC classes using OBIA and CNNs.
2. Evaluation of the OBIA methods and CNNs regarding their ability to automatically classify historical maps.
3. Creation of a vector data set of the Franciscean Cadastre, which can serve as a basis for further geographical analyses.

Besides the three main objectives, four secondary objectives were determined:

1. Creation of robust training and validation datasets that represent the characteristics of the digitized map source.
2. Finding the best segmentation parameters for the breakdown of the image into image objects, as part of the OBIA process.
3. Finding the best training parameters for the CNN.
4. Improvement of the base classification by implementing an expert-based reclassification section that incorporates additional conditions for selected classes.

Due to the availability of already digitized and georeferenced imagery of the Franciscean Cadastre, this study will omit parts of the preprocessing, focusing solely on the image analysis aspect of the map processing workflow, which will include the development of the CNN classification model and subsequent extraction of information from the historical map source.

2 Theoretical Principles of Map Processing

Before describing the exact methodology that was chosen for the classification of the Franciscean Cadastre, it is necessary to outline the basic principles of map processing or image analysis in the broad sense. The following chapter will introduce the most commonly used image analysis principles and methodologies used in digital map processing. Many of them are closely related to and often derived from methods used in computer vision, image processing, remote sensing, and digital cartography.

2.1 Map Processing Workflow

The overall workflow of digital map processing can be divided into three parts. The first part is centred around the initial scanning and digitization of the original map source and the georeferencing of the image file. The second part then focuses on the recognition and extraction of geospatial information from the map image using different techniques that will be explained in detail in the following sections. The third and last part is then focused on the vectorization of the extracted information by converting the raster dataset into a geographic vector dataset. A general sequence of the map processing steps is displayed in Figure 1.

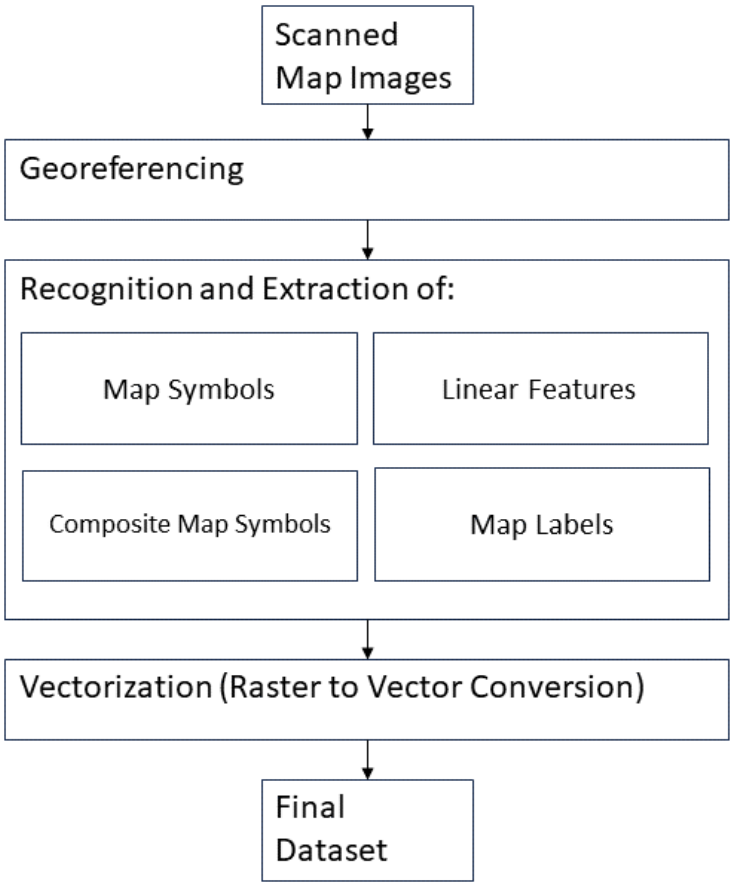


Figure 1: Example map processing workflow based on Chiang et al. (2014).

As computer-based map processing approaches aim to extract information from digital raster images, the typical map processing workflows begins after the map scanning process and the creation of the digital image. The digitized raster image represents the spectral characteristics of the original map in a machine-readable format. Because the scanned map image does not contain any information about the location of the map's content or its type of projection, the image is then georeferenced which provides missing spatial context to the data source. Alternatively, this data projection step can also be conducted at the end of the image processing. After this preprocessing of the image, the map image can be analysed and processed. The focus of all map processing methodologies usually lies in this section where the recognition and subsequent extraction of geographic information embedded within the map image is conducted. Workflows can be can be entirely automatic, require manual interventions during preprocessing/postprocessing, or be interactive (Chiang et al., 2014).

2.2 Image Preprocessing

Image preprocessing is often necessary to enhance the visual quality of the data source prior to conducting image analysis. Preprocessing encompasses a broad range of techniques and operations and may include noise reduction, pixel calibration, and standardisation to prepare the image for subsequent analysis. Preprocessing operations can be conducted on individual pixels, a local or neighbourhood environment or globally. Point operations target single pixels and do not consider the spatial organization of the image. Examples are contrast stretching, segmentation based on grey values, and histogram equalization. Local or neighbourhood operations use masks that include the spatial surrounding of the target pixel. Examples include convolution, with image smoothing or sharpening being the most know types of convolutions, spatial features detection, such as line, edge, and corner detections, and morphology. Global operations take the whole image into account and include frequency domain operations. A commonly used preprocessing method is *histogram equalization*. Image histograms provide quantitative information about the distribution of pixel values and are used for numerous spatial domain processing techniques. Histogram equalization is used to spread the histogram of pixel values more evenly using the full potential range. *Spatial image filtering* techniques use a moving filter mask to recompute pixel values of the image. Image smoothing and blurring filters build weighted averages of their surrounding pixels. Different types of smoothing filters exist, with low-pass and Gaussian filters being commonly used to reduce noise in an image. Contrary to image smoothing, edge detection and enhancement filters aim at highlighting local changes in colour. Several edge detection filters exist, with gradient, Laplacian, and wavelet transform being the most common categories (Mendoza and Lu, 2015).

In map processing, image preprocessing methods are commonly applied to improve the quality of the scanned map, prior to image analysis. Histogram stretching can be used to remove colour differences

between different map sheets, while smoothing and sharpening operations can be used to remove noise or highlight edges, depending on the properties of the map image.

2.3 Image Analysis

Centre of any computer-based image processing is the image analysis, which contains the detection and recognition of map features, their boundary delineation through segmentation, and the subsequent classification and extraction of features using various algorithms. When analysing map features, three different properties can be considered, colour, texture, and morphology. Colour analysis evaluates the spectral distribution of colours in the image, while texture analysis evaluates the spatial organisation of colours. Finally, morphology analysis describes geometric shape properties and size (Mendoza and Lu, 2015).

Over time, two basic strategies of image analysis have emerged, pixelwise image analysis and object-oriented image analysis. Both approaches differ fundamentally in how they handle information stored in the image. Pixelwise image analysis approaches use the absolute values of individual pixels to assign classes in the image. In contrast, object-oriented classification approaches, also called object-based image analysis, group statistically similar pixels in a preliminary segmentation step and then assign classes based on the geometric and spectral characteristics of the image segments (Gobbi et al., 2019).

2.3.1 Pixel-based Image Analysis

Pixel-based image analysis involves examining individual pixels in an image, focusing on the inherent information within each discrete pixel unit. This approach relies on the direct analysis of pixel values to extract features or perform basic image processing tasks. The fundamental concept is to treat pixels as standalone entities, making it suitable for tasks where local information is crucial. However, pixel-based analysis may face limitations in capturing complex spatial relationships and holistic features within the image (Blaschke and Strobl, 2001).

2.3.2 Object-Based Image Analysis

With the advent of (very) high-resolution imagery in the early 2000s, limitations the pixel-based image analysis approach became increasingly evident. Unlike with earlier, coarser imagery, landscape features in high-resolution imagery were now often represented by multiple pixels, creating new classification challenges. As pixel-based image analysis methods do not take into account neighbourhood information, salt and pepper effects will become increasingly common in high-resolution classifications (Blaschke et al., 2000) and therefore limit the classification accuracy of pixel-based image analysis approaches (Blaschke and Strobl, 2001). Blaschke et al. (2014) identified five key issues to be addressed in pixel-based image analysis approaches: objects, shape, texture, context and pattern, as well as semantics and knowledge integration (Blaschke et al., 2014).

A new approach to image analysis was required, that was able to look beyond the individual pixel for the extraction of spatial features. Based on these assessments Blaschke and Strobl (2001) argued for a radically different handling of entities, introducing the geographic concepts of neighbourhood, distance, and location to remote sensing (and image analysis). Prerequisite for the use of spatial concepts in image analysis is to move away from per-pixel analysis by aggregating information into so-called image objects, ideally representing spatially discrete objects in the imagery (Blaschke and Strobl, 2001). The solution was object-based image analysis (OBIA).

OBIA, also called geographic object-based image analysis (GEOBIA), is an image analysis approach that builds upon older remote sensing concepts such as image segmentation, edge-detection, feature extraction, and classification while providing a bridge to GIS by allowing for the integration of vector data and spatial analysis techniques into the classification workflow (Blaschke, 2010). For OBIA, a two-step approach is generally applied:

First, the image is partitioned into groups of spatially and spectrally homogeneous image objects, that represent real world objects on a defined scale using a region-based segmentation method. The identification of the correct scale is important, as meaningful objects only exist in a 'window of perception' (Marceau, 1999). Choosing the best fitting segmentation scale is highly depended on which objects are segmented. Finer scales might extract singular plants, while coarser scales might classify whole forests. It is also possible that different objects in an image appear different scales, requiring the introduction of sophisticated segmentation methods and hierarchy concepts (Blaschke, 2010).

In the second OBIA step, the image objects are then classified according to their geographical, geometrical, or statistical properties. Grouping homogeneous pixels together into image objects has several key advantages to image classification. On the one hand it prevents the salt and pepper effects present in pixel-based approaches, resulting in cleaner classification products. But more importantly, it allows the segments to have additional statistical information. This includes a variety of spectral (or in case of a map image colour) object descriptions such as minimum, maximum, and mean values, as well as additional spatial information for objects (Blaschke, 2010). This contextual information can help to improve classification results by either including more conditions in the initial base classification phase or by adding a follow-up reclassification using additional expert-based thresholds.

While OBIA has been successful in improving classification accuracies on high-resolution imagery, some challenges remain. Especially, identifying the best scale for the image segmentation remains challenging. In addition, fuzzy or smooth transitions present challenges to the segmentation algorithms creating issues with the correct delineation of class borders (Blaschke, 2010).

2.4 Image Segmentation

Image segmentation is a crucial step in every automated image analysis approach as seeks to identify, recognise, and delineate image objects, a task traditionally assigned to the human interpreter (Baatz and Schäpe, 2000). The discretization of the image is also a prerequisite for the subsequent classification. It is therefore important to achieve a robust segmentation result first, as any subsequent steps are highly dependent on the correct segmentation of the image (Mendoza and Lu, 2015). In OBIA segmentation is especially important as it produces the image objects that are used in the classification. The following section will provide an overview of the most important segmentation techniques used in image analysis.

2.4.1 Overview Image Segmentation Techniques

Over the time, several image segmentation techniques have been developed. Traditional segmentation methods can be grouped into pixel-based, edge-based, or region-based approaches (Blaschke et al., 2014, Ma et al., 2017). Even though numerous segmentation techniques exist, research is active and new methods are being developed (Ma et al., 2017). Chiang et al. (2014) identified five commonly used image segmentation techniques used for map processing:

Histogram thresholding is a pixel-based image analysis technique that divides the image into regions based on colour histogram of the image. Global and local maxima be set as criteria for class thresholds, if they correspond with the classes found in the map image. Colour histograms are usually created for each colour space dimension, but for the final assessment, individual results are usually combined. As histogram thresholding only evaluates colour histograms, spatial contiguity between classes is not explicitly accounted for, which can result in low accuracies for images without clear colour differences (Chiang et al., 2014).

Colour space clustering an unsupervised and multidimensional thresholding technique, that uses all colour dimensions available in an image and then attempts to identify statistically significant clusters in the colour space. For the clustering, K-means or the fuzzy set-based alternative C-means are frequently used. Partitions are done by measuring the distance (e.g., the Euclidean distance) in a colour space between the clusters centre and each observation. To improve clustering accuracy, nonparametric clustering methods, such as the mean-shift algorithms can be applied. As with histogram thresholding, spatial contiguity is not accounted for (Chiang et al., 2014).

Edge detection techniques try to detect contrasts between regions using local filtering operators. As they can only identify discontinuities in local colour values, they need to be applied in combination with other methods to be able to segment images. Edge detection methods generally show limited

success in images with poor contrasts or disconnected features or if significant noise is present (Chiang et al., 2014).

Artificial neural networks have also been used in segmentation tasks. They generally provide better results than other methods when dealing with blurred or mixed pixels, but require long training times and cannot be applied universally to all map types (Chiang et al., 2014).

Finally, *region-based* techniques divide images by grouping pixels into homogeneous groups by using region growing, region splitting, or region merging techniques. In difference to the previously named methods, region-based methods do take the spatial contiguity of homogeneous regions and connectivity between those regions into account, avoiding very small and scattered classes. Region-based techniques assume, that neighbouring pixels generally have similar values, which can be grouped. At the start of the process, seeds are selected from which surrounding pixels are tested for similarity to the seeds colour values and homogeneity (Chiang et al., 2014). The then algorithm tries to (iteratively) split and merge adjacent groups of pixels until a defined condition is met (Hossain and Chen, 2019). Regions can be grown, split, and merged according to the predefined conditions of homogeneity (Chiang et al., 2014).

2.4.2 Multiresolution Segmentation

Among the many region-based segmentation techniques that have been developed, multiresolution segmentation (MRS) has been particularly successful when combined with an OBIA workflow. As landscape features can have different sizes and spectral properties, they are often only partially captured by traditional region-based segmentation methods, which are only to a specific scale. This presents an issue for precise LULC classifications, as different objects may appear at different scales (Blaschke et al., 2000, Drăguţ et al., 2014).

MRS was first proposed by Baatz and Schäpe (2000) as a universal and reproducible method for image segmentation on multiple scales. The general approach resembles other region-based techniques, as it merges image objects based on a local homogeneity criterion. Each merging step is assigned with a merging cost, representing the degree of fitting. Merging stops if the least degree of fitting, the scale parameter, is exceeded. The merging process starts at the pixel level and then is executed bottom-up, merging neighbouring image objects if the homogeneity criterion is met. The homogeneity criterium itself is influenced by weighting spectral values of the pixels and the shape properties of the image objects (Baatz and Schäpe, 2000). The correct weighting of the segmentation parameters is depended on the aim of the classification process and the properties of the input image.

Analyses by Baatz and Schäpe (2000) came to the conclusion that while both colour and shape properties influence the segmentation results, colour appears to be a better criterion for optimal segmentations (Baatz and Schäpe, 2000).

A key challenge when using MRS algorithms lies in determining the optimal segmentation parameters. Especially the scale parameter, which controls the degree of heterogeneity within individual image objects, has a strong impact on the segmentation result. Smaller scale parameter values will generally lead to smaller segments with higher homogeneity, while a larger scale parameter allows for higher heterogeneity within image segments. But it is important to note that the scale parameter is not directly linked to a certain object size, which is why a trial-and-error phase is initially necessary for normal multiresolution segmentation to determine the scale parameter (Drăguţ et al., 2010).

To reduce the subjectivity inherent with the manual estimation the appropriate scale parameter, Drăguţ et al. (2010) and Drăguţ et al. (2014) developed the estimation of scale parameter (ESP) tool for the Trimble eCognition software, which automates this task by calculating the local variance (LV) of object heterogeneity within a scene and then iteratively generating image objects at multiple scales in a bottom-up approach until the thresholds in rates of change of LV (ROC-LV) are met. The ROC-LV measures the rate of change of the LV between each level. If peaks in the ROC-LV are assessed, it is assumed that a level has been reached that best represents image objects of equal homogeneity. The tool was tested on different image types and landscapes and proved successful in finding correct scale parameters in various different environments (Drăguţ et al., 2010).

The ESP tool can work with multiple layers, extending its usage to multispectral data and improving its applicability in OBIA workflows. For the computation of the ESP on multispectral images, the mean LV value is calculated for each image level. In addition, the ESP tool automatically creates three optimal levels (with logarithmic scale increments) in either a hierarchical or non-hierarchical approach to help with the selection of the best scale parameter (Drăguţ et al., 2014).

Despite the development of techniques for the optimization of the scale parameter, some challenges for region-based segmentation methods remain. In their meta-study Ma et al. (2017) identified five challenges: segmentation of linear objects, segmentation from low-level pixel grouping, multiscale segmentation, optimization of segmentation parameters, evaluation of segmentation results, and the continuing effort of achieving meaningful image objects from imagery (Ma et al., 2017).

2.5 Image Classification

Image classification, along with segmentation, is a crucial aspect of image analysis and map processing. Whilst segmentation divides an image into discrete classes, reducing the complexity of the original scene, the goal of classification is to assign pixels (or in case of OBIA image objects) to classes, giving

semantic meaning to parts of the image. Image classification can be conducted unsupervised or supervised. Supervised learning requires the definition of classes and the creation of labelled training datasets (Maxwell et al., 2018). While the creation preparation of supervised classifications is more resource intensive and time consuming, they usually provide higher overall accuracies. Because they are trained on labelled training data, model generalization and transferability to other imagery or map products can present a challenge.

An important task in setting up machine-learning classifiers, is the parameterization. Finding the optimal value for parameters is important and can significantly influence classification results. To prevent model overfitting and low generalization ability, training and validation datasets are split and the classifier is then evaluated on new data that it has not been trained on (Maxwell et al., 2018).

2.5.1 Overview Machine Learning Image Classifiers

For the classification task, different machine-learning algorithms are used, ranging from traditional methods to complex deep learning techniques. Traditionally, a variety of different machine-learning classifiers have been used for image classification (Ma et al., 2019). The following section will attempt to provide a short overview of commonly used classifiers:

Support Vector Machines (SVM) are a classification method aims at finding the optimal boundary between different classes by drawing a hyperplane in the feature space and simultaneously maximizing the margin, the distance between hyperplane and the nearest samples of each class. Non-linear relationships are addressed by SVMs by projecting the feature space to a higher dimension, better known as the kernel trick, which allows them to capture more complex patterns (Maxwell et al., 2018). Advantages of SVMs are the comparatively low number of training samples required for good classification results and their ability to handle high dimensional data, such as multispectral imagery (Ma et al., 2019).

Decision Tree (DT) classifications are recursively splitting the input data based on thresholds. Repeated splitting creates tree-like branches that represent different paths, with the ends of the branches in the classification trees representing the final classes. If-then rules applied during the decision tree classification make DTs a simple classifier that produces fast results but is prone to non-optimal solutions and overfitting (Maxwell et al., 2018).

Random Forest (RF) is an ensemble classifier that uses multiple DTs to overcome the weaknesses of a single DT. Final classes are calculated by taking the majority vote of all trees. It thereby reduces the issues connected with singular DTs (Maxwell et al., 2018). Compared to SVMs, RFs are easier to use and generally produce high accuracy classification results (Ma et al., 2019).

The *k-nearest neighbour* (K-NN) algorithm is unlike other classifiers, as it is not trained to produce a classification model. Each unknown sample is instead directly compared against the original training data and then assigned to the most common class that is nearest in the feature space (Maxwell et al., 2018).

Lastly, *Artificial Neural Networks* (ANN) emulate biological neural networks by organizing neurons in interconnected layers. Between the input and output layer are several hidden layers that connect the neurons or nodes to adjacent layers. Each connection node receives signals from the previous node, which are computed by a non-linear function, called the activation function. The signal is then sent to the next nodes, where the process is repeated. Each node also has a weight that influences the strength of the signal, thereby controlling the learning process. Nodes in ANNs are hierarchically structured, where each layer represents a step in the processing with different transformations. Signals travel from the input layer to the final output layer via the intermediate hidden layers. If an ANN has two or more hidden layers, it is considered to be a deep neural network (Ma et al., 2019).

Although the concepts of ANNs are not new, their application in image analysis was limited for a long time, as they are considered to be slow to train, can produce non-optimal classifications, and are prone to overfitting. Additionally, the many user-determined parameters can make training of deeper networks challenging (Maxwell et al., 2018). These challenges combined with hardware limitations did restrict the practical application of ANNs for a long time and led to a focus on other classifiers. While many challenges persist, continuous advances in computer hardware, especially in graphics processing units (GPUs), have led to renewed interest and a surge in scientific publications since 2014. Because ANNs have been very successful in extracting high-level feature information, their application to high-resolution imagery was evident for a long time. Recent research has shown that ANNs have been able to outperform other traditional classifiers, such as RF, SVM and DTs, in supervised object-based image classification where none of them exceeded median accuracies of over 90% (Ma et al., 2019).

2.5.2 Convolutional Neural Networks

Convolutional Neural Networks (CNN) are one of the most widely used type of deep artificial neural network. The first modern CNNs were introduced by LeCun et al. (1989) where they were used to regcognize handwritten digits on letters of the U.S. Postal Service. The approach proved to be successful, as the model was able to learn the morphology and texture of different digits, providing good classification results and low overall error rates. It was also able to correctly classify atypical data, which made it more flexible than other approaches (LeCun et al., 1989). Subsequent research improved the model, but further major progress was limited by the high hardware requirements and the general lack of training data (Gu et al., 2018).

While CNNs were originally designed to process multiple array data sets, they were soon found to be well suited to process multi-band imagery, making them capable for colour image analysis. Since then, they have been successfully applied for the analysis of high-resolution remote sensing imagery with spatial resolutions of less than 10 m (Ma et al., 2019), as well as historic maps (O'Hara et al., 2024). Research has shown that the design of CNNs, such as the number of feature maps and hidden layers, has direct impact on its recognition capabilities. As a rule of thumb: Shallow CNNs with few hidden layers are better suited to learn general features and detect edges, while deep CNNs with many hidden layers can recognize more specific features (Chiang et al., 2014), such as specific types of objects or map symbols.

CNNs generally consist of three different types of layers: convolutional layers, pooling layers, and fully connected layers (see Figure 2). Like other ANNs, they are structured hierarchically, starting with the input layer. Then, convolutional layers and pooling layers are concatenated. The processing ends with a fully connected layer at the end.

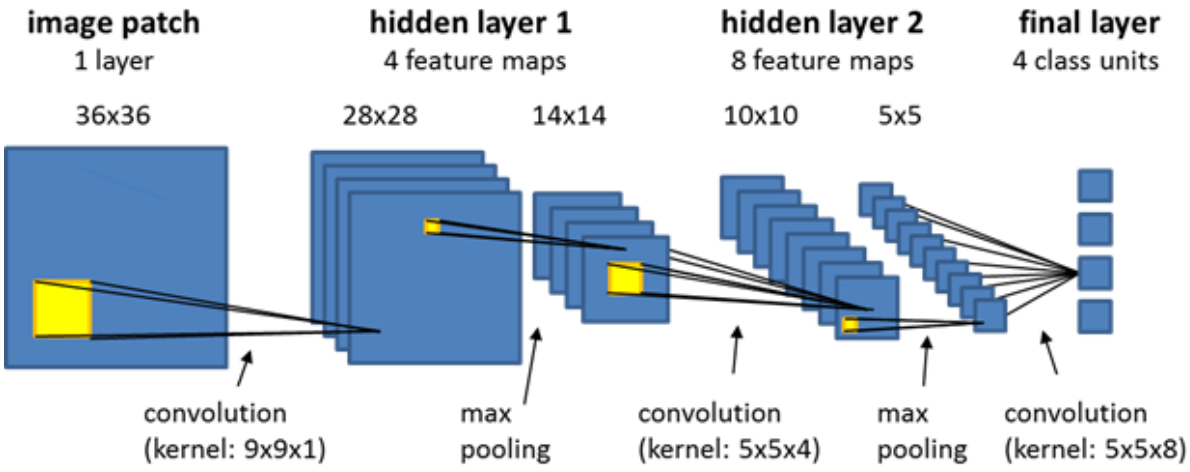


Figure 2: Hierarchical structure of a simple CNN with two hidden layers (Trimble, 2024b).

The convolutional layer is composed of multiple convolution kernels, that compute different feature maps. Each neuron of a feature map is connected to other neighbouring neurons in the previous layer, creating a field of perception for the neuron in the previous layer. A new feature map is processed by convolving the input layer with a learned kernel and then applying an element-wise nonlinear activation function on the convolved results, typically the sigmoid function (Gu et al., 2018).

The primary objective of the pooling layer is to establish shift-invariance by diminishing the resolution of the feature maps. As a pooling layer is usually positioned after a convolutional layer, each feature map of the pooling layer corresponds to the preceding feature map of the convolutional layer. Pooling operations are typically divided into average and max pooling. In a typical CNN, the kernels of the first

convolutional layer are designed to detect the most basic features such as edges and curves, while kernels in following layers aim to detect more abstract features, such as different types of digits or vegetation classes. The number of convolutional and pooling layers in a CNN therefore influences the ability to extract high-level feature representations. After several convolutional and pooling layers, a fully connected layer may take all neurons of the last layer and connect them to generate global semantic information. For classification tasks, a softmax operator or a SVM may be used (Gu et al., 2018).

A key challenge in the creation of a CNN is the definition of weights and biases, which heavily influence classification results. Optimizing these parameters is therefore an important step in every training process. To be able to assess the level of optimization and update the model's parameters, a loss function is usually calculated using stochastic gradient decent (Gu et al., 2018).

Sample data augmentation sets to increase the amount of training samples using a broad variety of data transformations. Common augmentation methods in image analysis are geometric transformations, such as mirroring, rotating, and shifting, as well as various types of photometric transformations. As data augmentation does not require the manual selection of samples, it is especially useful, if there are no large quantities of training data available and the production of new samples is too time consuming. To prevent overfitting of the data, the training process can be terminated using early stopping, a method which halts further optimization if certain conditions are met. Training can be stopped either after a certain number of epochs has been passed or if a predefined training error is reached (Gu et al., 2018). Another way to improve a CNN model, proposed by Sergey and Christian (2015), is to use batch normalization. The method normalizes each training batch within the CNN and allows the model to use higher learning rates, reducing the overall training steps and training time necessary to reach sufficient classification accuracies (Sergey and Christian, 2015).

In recent years, CNNs have been used for a variety of different tasks, including image classification, object detection, object tracking, text detection and recognition, as well as scene labelling. While CNNs provide many advantages to traditional machine-learning techniques, the biggest barrier to their broad application still exists: Training CNNs and adjusting the models hyperparameters, such as learning rate, kernel size, convolutional filters, and number of layers, requires considerate knowledge and experience (Gu et al., 2018).

2.6 Rule-based Classification

While machine-learning classifiers are a popular approach to image classification, alternatives exist. One alternative, that is also frequently used, is rule-based classification (also called expert-based classification). Rule-based classification is conducted by using expert rules and thresholds through the implementation of simple if-else clauses to (re-)classify pixels or image objects (Shivakumar and Nagaraja, 2023). Combining rule-based classification with supervised classification approaches can refine the overall classification process by adding further conditions that improve the accuracy for certain classes. It can also be used to retrospectively correct some issues related to machine-learning based classifications, such as classes that were not adequately covered by the provided training data. Classes can be split or merged through thresholds after the main classification has been conducted, creating more detailed classification results. For instance, it could be used to differentiate built-up structures into urban and rural based on spatial proximity. Rule-based classification is therefore frequently part of the OBIA framework where it helps to enhance classifications.

2.7 Accuracy Assessment

Once the initial classification results have been obtained, it is necessary to assess the model's performance to evaluate the overall effectiveness of the chosen approach. Accuracy assessment is a fundamental requirement in image analysis and map processing to evaluate the model's generalization ability and transferability to other datasets. For the evaluation, multiple criteria exist: Overall accuracy, precision, recall, and the F1 score (Kersapati and Grau-Bové, 2023).

The criteria are calculated using the following equations:

$$Accuracy = \left(\frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FP} + \mathrm{FN} + \mathrm{TN}}\right)$$

$$Precision = \left(\frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}\right)$$

$$Recall = \left(\frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FN}}\right)$$

$$F1\ Score = \left(\frac{2 \text{ x Recall x Precision}}{\mathrm{TP} + \mathrm{FP} + \mathrm{FN} + \mathrm{TN}}\right)$$

All criteria are calculated using the measured numbers of true positives (TP), false negatives (FN), false positives (FP), and true negatives (TN), obtained from the statistical comparison between the classification result and reference data (Kersapati and Grau-Bové, 2023). The creation of confusion matrices helps to visualize errors of omission and errors of commission in a matrix. The matrix contains all number of sample units and their classification class. Samples can be points, cluster of points, or polygons. Based on the sample distribution and differences between reference and classification data, several classification criteria can be calculated (Congalton, 1991).

Besides these statistics, Cohen's Kappa has been used to assess classification results. Kappa is calculated using the following equation:

$$Kappa = \frac{\text{observed agreement} - \text{exspected agreement}}{1 - \text{expected agreement}} = 1 - \frac{1 - \text{observed agreement}}{1 - \text{expected agreement}}$$

The Kappa measure allows comparisons between the obtained classification results and a classification using random guesses. Kappa results are always less or equal to 1, where negative values indicate no improvements to a random classification (Kersapati and Grau-Bové, 2023).

When performing accuracy assessments, several things need to be considered to achieve proper results. Congalton (1991) found five factors that can influence accuracy assessment results: ground data collection, classification scheme, spatial autocorrelation, sample size, and sampling scheme. Regarding the sampling size, they set the minimum number of samples per class to 50, but the exact count depends on the size of the area and the complexity of the scene. Another topic of scientific discussion is the sampling scheme, as it can heavily influence the accuracy assessment outcome. Opinions vary from simple random sampling to stratified systematic unaligned sampling (Congalton, 1991).

3 The Franciscean Cadastre

The Franciscean Cadastre (in German: „Franziszeischer Kataster“, alternatively „Franciszäischer Kataster“, was the first comprehensive and geometrically accurate mapping of parcels and land use in the Austrian Empire. Surveying was conducted between 1817 and 1861 on behalf of Emperor Francis I. after the end of the Napoleonic Wars and the restoration of the Empire and its territories as a result of the Congress of Vienna in 1815. The cadastre was created in several stages, with the province of Salzburg being surveyed between 1823 to 1830, while the evaluation and treatment of all complaints by landowners lasted until 1844. Goal of the great survey was to create a central detailed cadastral inventory which could serve as the basis for a long-planned reform in taxation. In difference to previous approaches, land taxes should be adjusted not only to yield and price ratios, but also to production costs and thus to the net agricultural income. Prerequisite for this tax estimation was an accurate, parcel-specific land survey of cultivatable land, which was realized by the new survey. The second reason for the creation of the cadastre besides the aforementioned tax reform was that such a cadastre would allow for the direct access to the resources of the empire’s inhabitants. It would support the creation of a unified fiscal space, an important step towards the formation of modern nation states. Already shortly after the first surveying and mapping was finished, the inventory proved to be an important tool for planning, as it contained detailed information about the Empire’s landscapes and economic life (Gebhart, 2011).

The Franciscean Cadastre can be divided into several parts. The best-known part, which is at the same time the central part of the cadastre is the detailed cadastral plan, containing the parcel, building, and land-use information with boundary descriptions. Besides that, a land and building parcel protocol was created, an alphabetically ordered directory of land parcel and building owners, as well as a tax estimation operation, which served to determine the value of the income and production expenditure. For the documentation of demography, information of population, house, and household sizes within the cadastral municipalities were recorded, providing information about the social structure of society. The Franciscean Cadastre thus represents one of the most comprehensive and valuable historical sources of the 19th century for the area of the Habsburg Monarchy, providing a detailed insight (see Figure 3) into the agricultural economy on the eve of industrialization for large parts of central Europe and the Balkans (Bauer, 2017).

Figure 3: Detail from the Sattler panorama by the Austrian painter Johann Michael Sattler (photo by WOKRIE (2015)). The painting captures the pre-industrial landscape of Salzburg during the creation of the Franciscean Cadastre.

As the Franciscean Cadastre was created using standardized methods, it is particularly well suited for quantitative landscape analysis. Therefore, it has been thoroughly analysed in research by historians and geographers to assess changes to the cultural landscape (Gabrovec et al., 2019, Dolejš and Forejt, 2019). For example, Bicik et al. (2019) analysed the long-term land-use change in Czechia and Slovenia by using the Franciscean Cadastre as reference for 19th century land-use, and then compared both developments over course of the 20th century. Kupkova et al. (2019) analysed the development of land-use and land cover in the Czech border regions after the expulsion of the ethnic Germans after the Second World War. The pre-1945 land-use data was in parts based on the Franciscean Cadastre. Foski and Lamovsek (2019) created landscape metrics indices for the monitoring of land-use change using the Franciscean Cadastre as point of reference. The impact of land-use change on ecosystem services was analysed by Ribeiro and Hribar (2019) who used data of the Franciscean Cadastre as supporting data source in their research. Because the Franciscean Cadastre was mapped on a fine scale of 1:2880 for rural and 1:1440 for urban areas, providing detailed information about buildings, roads, and landscape (Land Salzburg, 2012), it allows for extensive LULC change analysis on local scales (Gabrovec and Kumer, 2019). Further studies that used the Franciscean Cadastre as a source of information have been summarized in Dolejš and Forejt (2019) meta-analysis about use of the Franciscean Cadastre in historical research.

4 Extracting Information from the Franciscean Cadastre

The Franciscean Cadastre contains rich spatial information about large parts of Austria and the Balkans before the advent of widespread industrialization. Harnessing this information has been a goal of many studies in the past, but the quantitative use in GIS requires the digitization and vectorization of the original data source. As parts of the cadastre have already been scanned and georeferenced, this study is focused on the automatic classification and vectorization of parts of the Franciscean Cadastre instead. This chapter describes the workflow used to extract geographic information from the Franciscean Cadastre and transform it into vectorized geodata.

The scanned data of the Franciscean Cadastre was kindly provided by the Department 7/06 – spatial data infrastructure, Land Salzburg. The data was acquired in the form of digitized and already georeferenced images in the GeoTIFF file format, projected in MGI Austria GK M31, the regional coordinate reference system. In total, two regions were requested, each covering about 20 km².

The first region, used for the training and testing of the classification model, is located around the city of Salzburg (see Figure 4). This section was chosen, because it includes significant urban areas and different types of roads as well as LULC. In detail, it contains the historic city centre, the city districts Mülln and Nonntal at the gates of the city, and the then independent village of Maxglan in the west. This diversity in geographic features provides an excellent testing ground for classification and recognition algorithms aiming to identify different-sized features within the map. The inclusion of Salzburg in the research is also motivated by the heterogeneity in the quality of data sources. Some map sheets within the area of the digitized cadastre are discoloured, introducing challenges to the classifier. This aspect adds a layer of realism to the study, mirroring the conditions often encountered in the use of historic map products and map processing. In addition to the algorithmic challenges, the study area of Salzburg was also chosen because of the city itself. Its history has been subject to various research for a long time, but especially historic geographic data is still lacking, obstructing further research in this area.

The second study area was set in the High Tauern mountain range between the villages of Taxenbach and Rauris in the south of the province of Salzburg (see Figure 5). The region provides a different cultural landscape with a predominant share of alpine pastures and coniferous forests and fewer roads and buildings. In addition, the region contains an additional map processing challenge in form of severe yellowing that is found in the original paper map sheets and has been transferred to the digitized map image. The second study area was mainly chosen to test the transferability and robustness of the classification approach in a different map section to assess the capabilities for potential large scale map vectorization works in the future.

Figure 4: Main study area around the city of Salzburg.

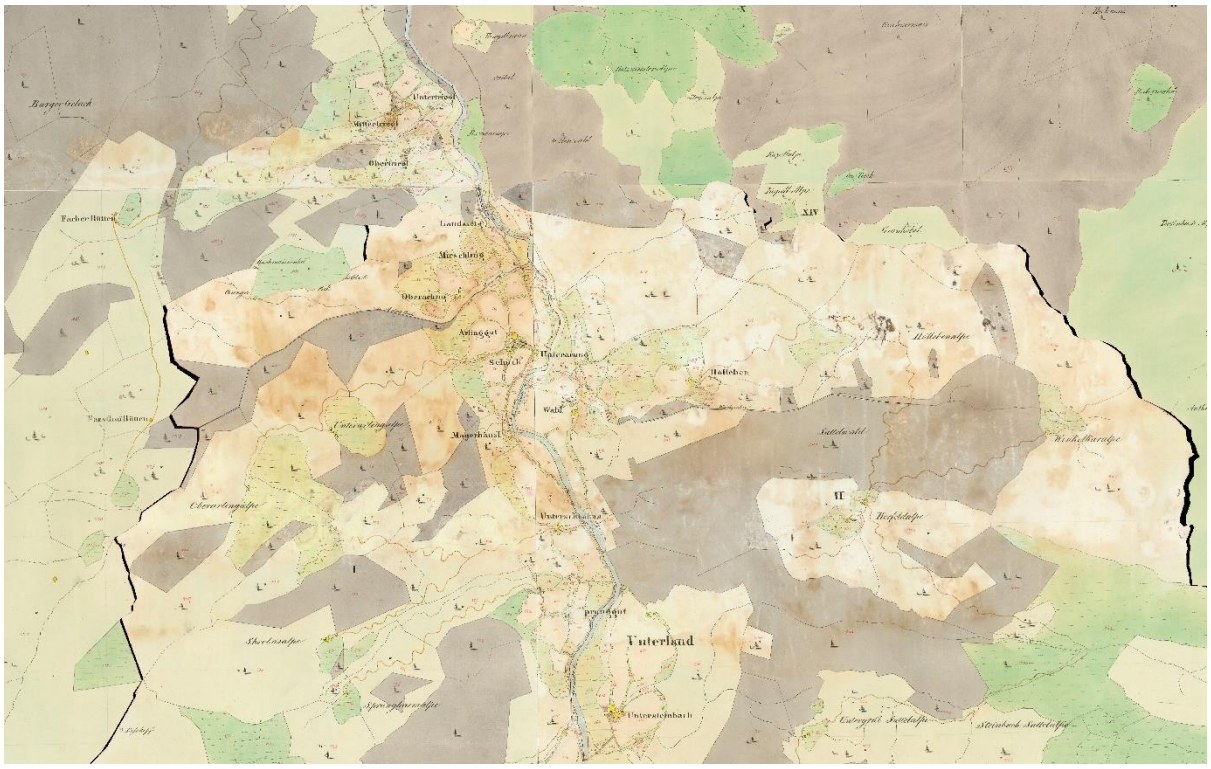


Figure 5: Second study area between Taxenbach and Rauris.

4.1 Software

Several commercial image analysis products are available, complemented by various free open source options, that are capable to automatically extract spatial information from images. For this study, the commercial product Trimble eCognition 10.3 was used, as it provides an extended palette of features for (object-based) image analysis including multiple algorithms for geographic information extraction and data fusion and the ability to combine remote sensing workflows and GIS analysis (Trimble, 2024c). For the creation of training, validation, and testing data, as well as the accuracy assessment, Esri ArcGIS Pro 3.2.1 was used.

4.2 Workflow

The workflow developed for the classification of the Franciscean Cadastre follows the general image analysis methodology described in chapter 2 Theoretical Principles of Map Processing'. The first part of the analysis is focused on the digitization and pre-processing of the map source. For the training and validation of the classification models, multiple training datasets have been created. Training sets were created for the LULC classes and for the selected map symbols.

After preparing the input datasets, the working environment was then shifted to eCognition. In eCognition the main image analysis was conducted which did include the LULC classification and recognition of map symbols. Within the eCognition workspace the work was split into four projects representing the four map sections (grouped into two regions) received form the State of Salzburg. The training of the CNN classification model was conducted on the map image of eastern Salzburg using the training samples created in the preprocessing. Then, the model was applied to the map section of western Salzburg. Both map sections in the High Tauern region, were disregarded and only used for the final transferability assessment. As outlined in section 2.3.2 Object-Based Image Analysis', the LULC image analysis consists of two main parts, the segmentation of the map into discrete homogeneous regions, and the map classification based on the class probability heatmaps generated by the CNN. To enhance the classification results and remove unwanted cartographic symbols and text, an expert-based reclassification was also implemented before the results were converted into a vector dataset and exported. In a final step, the classification results for both study areas were evaluated through an accuracy assessment. The confusion matrix of the accuracy assessment was created in ArcGIS Pro. An overview of the implemented workflow can be seen in Figure 6.

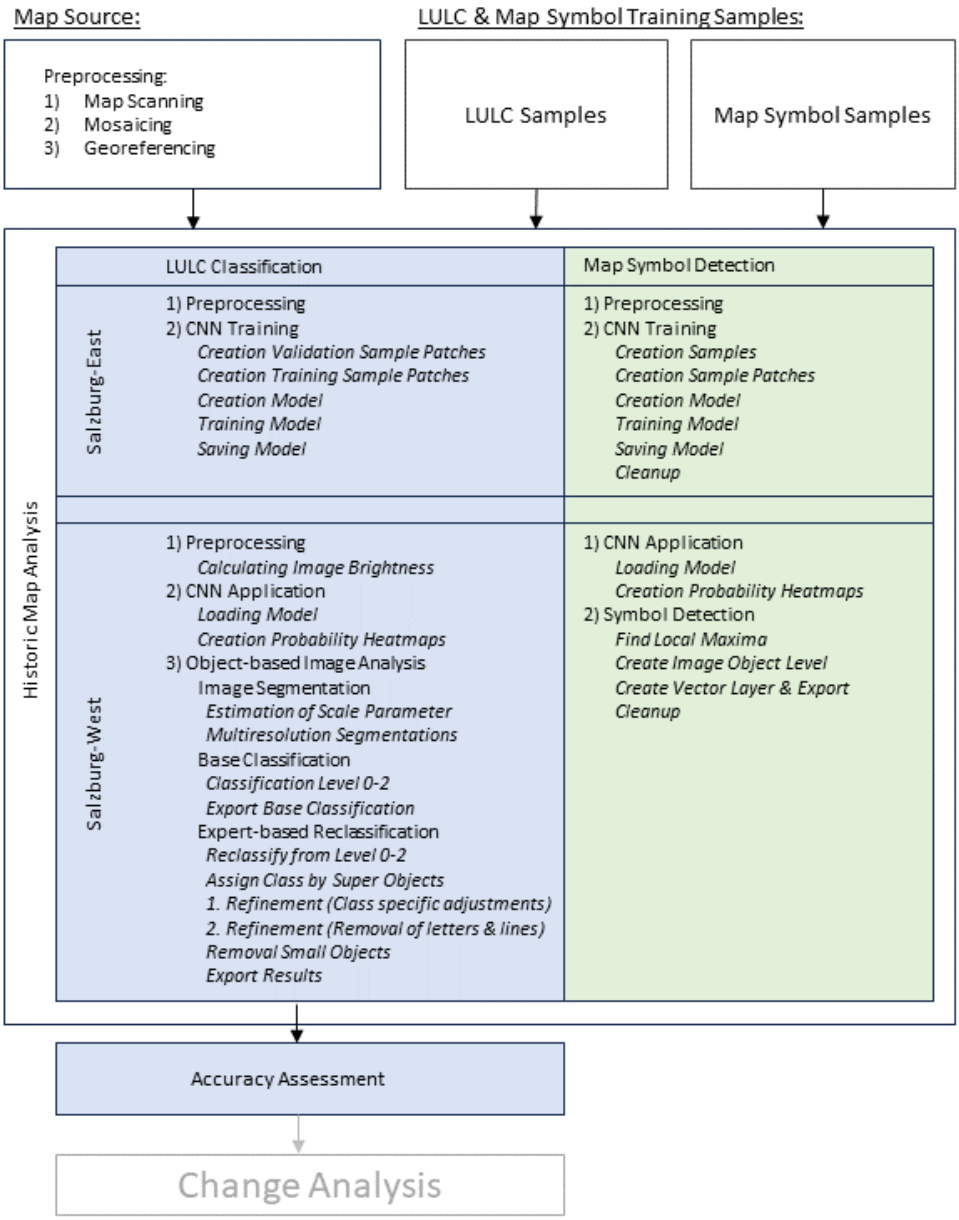


Figure 6: Overview of the General Workflow.

The following chapters will describe in detail the algorithms and rulesets used for the processing of the Franciscean Cadastre, including parameters used for segmentation and classification.

4.2.1 Preprocessing

Prerequisite for any supervised classification using machine-learning algorithms is the creation of labelled samples. Samples are usually divided into training and validation sets by withholding some samples from the training process (Maxwell et al., 2018). The total quantity and the ratio between training and validation samples differs and is subject to scientific discussions, but commonly rations of 90:10 or 80:20 for training and validation samples are used (O'Hara et al., 2024). For this project, two

different point-sample training datasets were created using ArcGIS Pro, one dataset for the training of the map symbol recognition algorithm, and one for the training of the LULC classification algorithm.

For the map symbol recognition, 576 unique training samples were taken within the 'Salzburg-East' training area. Map symbol types were restricted to different tree and shrubbery types mainly, to limit the workload of the project. A total of six classes were defined, including oval and spiky coniferous symbols, oval and round deciduous symbols, meadows, and small shrubbery. Because map symbols are very unevenly distributed in the study area, counts per sample differed greatly as seen in Table 1 below.

Table 1: Sample count for the different map symbol types before the sample augmentation. Sample counts of coniferous trees in the training area was low, necessitating data augmentation strategies.

Class Name	Training Sample Count
Coniferous oval	20
Coniferous spiky	19
Deciduous oval	117
Deciduous round	257
Meadows	94
Shrubbery small	69

The LULC classification model requires the creation of training sample points as well. 200 training samples were initially created. From this batch, a small number was split and used to validate the model during the training process. For LULC classification training sample dataset 14 classes were created (see Table 2 below), based on the main LULC types documented in the Franciscean Cadastre.

Table 2: Sample Count for the LULC types before the sample augmentation. NoData has a lower training sample count, as the class is not very present in the region and the spectral complexity of the class with mostly black pixels is low.

Class Name	Training Sample Count	Validation Sample Count
Agriculture	160	20
Black Line (BLine)	160	20
Brown Path (BPath)	160	20
Forest	160	20
Garden	160	20
Meadow	160	20
NoData	100	20
Other	160	20
Red Line (RLine)	160	20
Red Path (RPath)	160	20
Red Urban (RUrban)	160	20
Water	160	20
Yellow Path (YPath)	160	20
Yellow Urban (YUrban)	160	20

The different classes in the LULC classification dataset can be categorized into three distinct groups. The first group contains cartographic classes that do not directly represent LULC but cover significant parts of the image. They must be included in the classification model because they represent significant parts of the map but will be eventually removed at a later stage in the classification. This group is comprised of black line, red line, and no data classes. The second group is made up of red urban and yellow urban areas which represent built-up areas, and brown path, red path, and yellow path, which are different types of road classes. The remaining classes are part of the core LULC group, comprised of agriculture, forest, garden, meadow, water, and others, that make up most of the land cover.

While the original Franciscean Cadastre contains several more classes and subclasses, only the ones with distinct colouring can be considered. Further differentiation and subdivisions into different types of gardens, agriculture, and forests in the original maps were made using map symbols including different types of trees, crops, abbreviations, and icons (see Figure 7). To be able to extract those classes, map symbol recognition is required, which was addressed in the second part of the classification methodology.

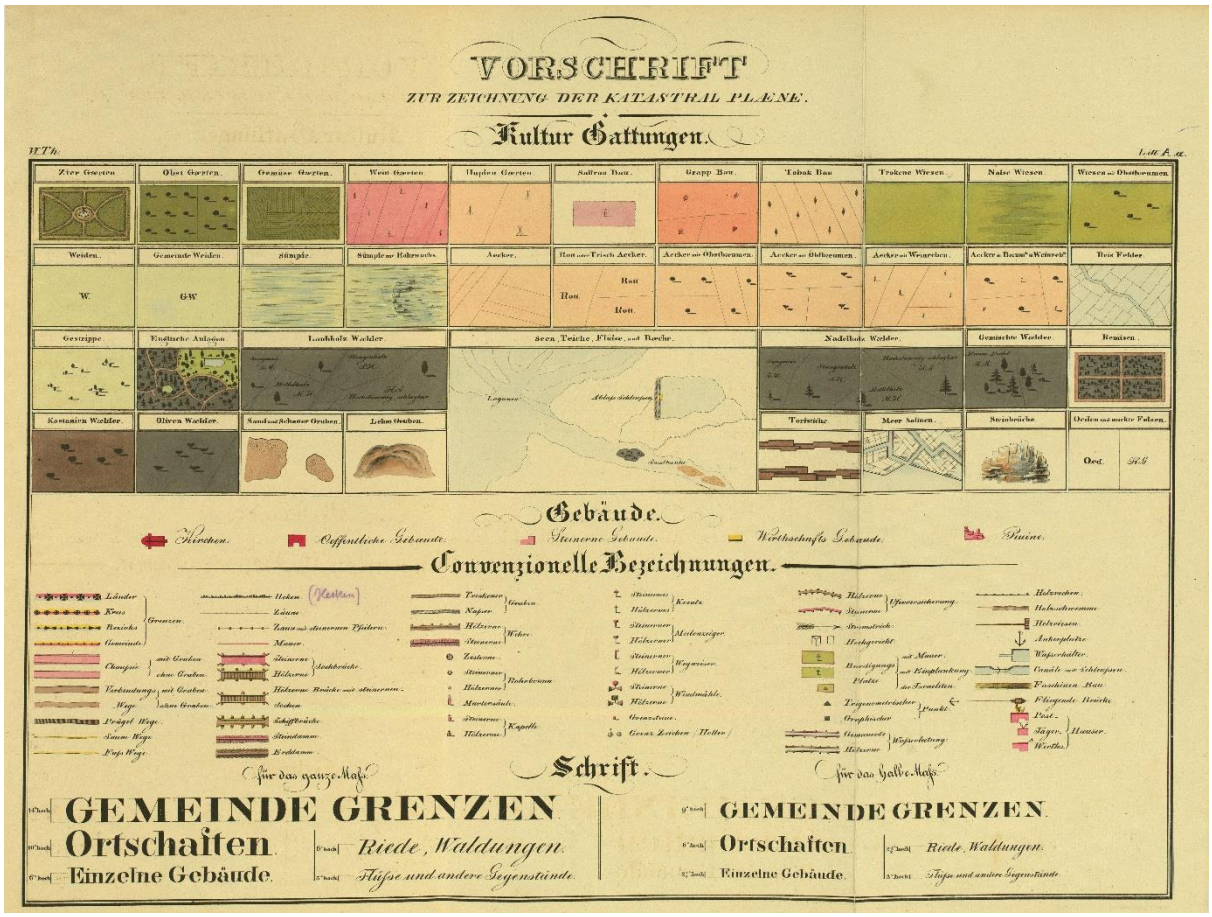


Figure 7: Legend of the Franciscean Cadastre from 1824 (Wikimedia Commons, 2014).

Besides LULC classification samples taken in the 'Salzburg-East' training area, additional accuracy assessment point samples were collected in the 'Salzburg-West' and High Tauern study areas. For the generation of sample points, a stratified random approach with a target of 500 points was chosen to better reflect the distribution of classes in the study areas. Samples were generated in ArcGIS Pro using the 'Create Accuracy Assessment Points' geoprocessing tool. The number of points generated per class varies based on their share of land cover (see Table 3).

Table 3: Accuracy assessment sample count for the LULC classification in the 'Salzburg-West' and High Tauern areas.

Class Name	AA Samples - Salzburg-West	AA Samples - High Tauern
Agriculture	268	
Brown Path (BPath)	7	9
Forest	29	84
Garden	16	4
Meadow	123	412
Other	21	2
Red Path (RPath)	9	5
Red Urban (RUrban)	9	
Water	29	18
Yellow Path (YPath)	1	1
Yellow Urban (YUrban)	8	9
	526 in total	544 in total

As the samples were generated based on a stratified random approach, they are not distributed equally between the different classes. For example, yellow paths were only sampled once for each location, and in the High Tauern area agriculture and red urban classes were missing entirely.

4.2.2 Map Symbol Recognition

As a cadastral map, the Franciscean Cadastre contains a great variety of information about the use of land parcels. Information is conveyed using different colours, textures, and symbols. The map symbols support the classification of different LULC classes, especially vegetation, providing detailed information about the exact type of forests, gardens, and meadows. Information about road types is also enhanced by the map symbology, where the existence of trees points to avenues.

Six different vegetation symbols were identified for the image analysis. The classes were grouped into coniferous trees (containing an oval and a spiky type), deciduous trees (containing an oval and a round type), small shrubbery, and meadows. The meadows were symbolized by the letter *m*, indicating meadows in areas where detailed colouration would not be applicable. All map symbols were coloured in a dark brown to black colour tone. For the recognition and extraction of the different map symbols, a deep learning approach was chosen over template matching, as the map symbols are handwritten and come in variations which could provide issues with simpler template matching approaches. The

general workflow for the map symbol recognition is based on a tutorial for object recognition in eCognition by Trimble (Trimble, n.d.), with adjustments made to be able to classify multiple symbol types simultaneously.

4.2.2.1 Training

The training process started with the import of the map symbol training samples and the creation of a buffer around the training points with a 7-pixel offset. Afterwards, the target symbols were created using a vector-based segmentation on the buffered samples layer, resulting in a new level called Level_S3. On Level_S3, the segments were classified based on the class name in the training samples dataset. To generate non-target segments, a new brightness layer was calculated by using the algorithm '*layer arithmetics'* and the formula: $\frac{(red + green + blue)}{3}$. The brightness layer was then smoothed with a Gaussian filter with a kernel size of 5, followed by a multi-threshold segmentation that was applied to detect and segment unclassified areas below a threshold of 140. The darker areas which did not intersect buffered training points were then classified as non-target. After creating sample and non-sample segments, all class segments were split into image segments using chessboard segmentation.

After completing the pre-processing phase, the 'generate labelled sample patches' ruleset was used to generate sample patches for the training. A total of 2000 samples were generated for the classes 'Coniferous_oval' and 'Coniferous_spiky', with a sample patch size of 32 and using the red, green, and blue image layers. As the initial sample collection only generated insufficient numbers of samples for these two classes, sample augmentation techniques were used to increase the sample size artificially to 8000 per class to enhance the CNN algorithm's robustness. The input sample size was multiplied by four using sample rotation. Samples were rotated between zero and four degrees and only a slight rotation was permitted as symbols are typically not rotated on maps, and higher rotations could result in false detections. Horizontal or vertical flips were also disabled, as symbols are not symmetrical and cannot be mirrored. Sample patch normalization remained disabled. The other four sample classes 'Deciduous_oval', 'Seciduous_round', 'Shrubbery_small', and 'Meadows' 8000 sample patches were generated using no sample augmentation, as their input sample size was higher than for the coniferous samples. For the non-target sample class, which mostly contained letters and lines, 8000 sample patches were generated using additional sample augmentation with a 4x multiplier using utilizing horizontal and vertical flips to get to a final sample patch count of 31920. More samples were generated for the non-target class because it is the dominant class on every map.

After generating sample patches for each of the seven classes, the CNN model was created and adjusted for symbol detection. The model was set to use only one hidden layer with a kernel size of 13

and 40 feature maps. Batch normalization and max pooling were enabled. The model was then trained for one epoch on the shuffled training data. For the training, the learning rate was set to a value of 0.0015 at 2500 training steps with a batch size of 32 to ensure that the model would see each sample exactly once. Once the model was trained, it was saved, and all image layers and object levels were deleted.

4.2.2.2 Symbol Extraction

After switching the project to the 'Salzburg-West' study area, the trained model was applied to generate the probability heatmaps. The generated heatmaps were then smoothened using a Gaussian filter with a kernel size of seven, to reduce noise. Then, local maxima in the heatmaps were detected by dilating the classes using the '*pixel filter 2D*' ruleset. For this ruleset, the morphology filter option was chosen. The dilation operation was iterated nine times. Target locations for the symbols were calculated using the '*layer arithmetics'* ruleset and a threshold of 0.95 likelihood using the following equation:

$$targetLOC = \big(heatmap\_localMAX = heatmap \& (heatmap\_localMAX > 0.95)\big) * heatmap$$

To increase the area of the target locations, the dilating morphology filter was applied a second time with three iterations. Afterwards, a new image object level called 'Level_Symbols' was created using multi-threshold segmentation, which created classified segments around the target locations with a greater than 95% likelihood of being certain map symbols. In a last step, the segmented areas were converted to points by using the '*convert image objects to vector objects'* ruleset. The location of the individual points was set to the centre of gravity of the image objects. Then, the points dataset was exported as shapefile with the '*export vector layer'* algorithm.

4.2.3 LULC Classification

4.2.3.1 Training

Like the training of the map symbol recognition model, the training of the LULC classification model was conducted in the 'Salzburg-East' study area. After importing the training and validation samples for the LULC classification, all sample points were buffered using a round buffer with a radius of nine pixels. Through a vector-based segmentation, the buffered samples converted to image segments in a new image object level. On the new level, the image objects were classified based on the class attributes of the overlapping samples. To prevent training issues with samples close to the image border, all samples with a distance less or equal to 24 pixels to the border were unclassified. Afterwards, sample patches were generated using the '*generate labelled sample patches'* algorithm. In difference to the training of the map symbol recognition model, no non-target classes were defined, as the whole image would be classified with the previously defined LULC classes.

To increase the number of samples and therefore the robustness of the algorithm, data augmentation methods were again implemented. Image rotation was introduced using 12 different rotations. Additionally, samples were flipped horizontally, and vertically to increase the sample variety. The sample patch size was set to 48x48 pixels. For the training of the LULC classification model 168000 training samples and 33600 validation samples were generated, covering all 14 classes.

After generating the sample patches, the neural network was created. The CNN was designed with two hidden layers. The first hidden layer had a kernel size of 3 and 12 feature maps and used max pooling, while the second hidden layer had a kernel size of 3 and 24 feature maps. In difference to the first hidden layer, the second hidden layer would have no max pooling. Batch normalization was activated. The training was conducted with a learning rate 0.0006 and 5250 training steps with a batch size of 32. The best patch size for the classification of historic maps is the topic of scientific discussions. For this approach, a relatively small patch size was chosen because the map contains very fine lines and symbols. But other studies such as O'Hara et al. (2024) have also achieved success with larger samples patch sizes up to 512x512. Parallel to the training of the algorithm, the model was validated with the the validation dataset. Training and validation steps were adjusted so that all samples would be used during one run of the algorithm. Each epoch, training loss and validation loss, as well as training accuracy and validation accuracy were written into an array and before each new training round all samples, including the validation samples, were shuffled to prevent the algorithm from recognizing patterns in the samples.

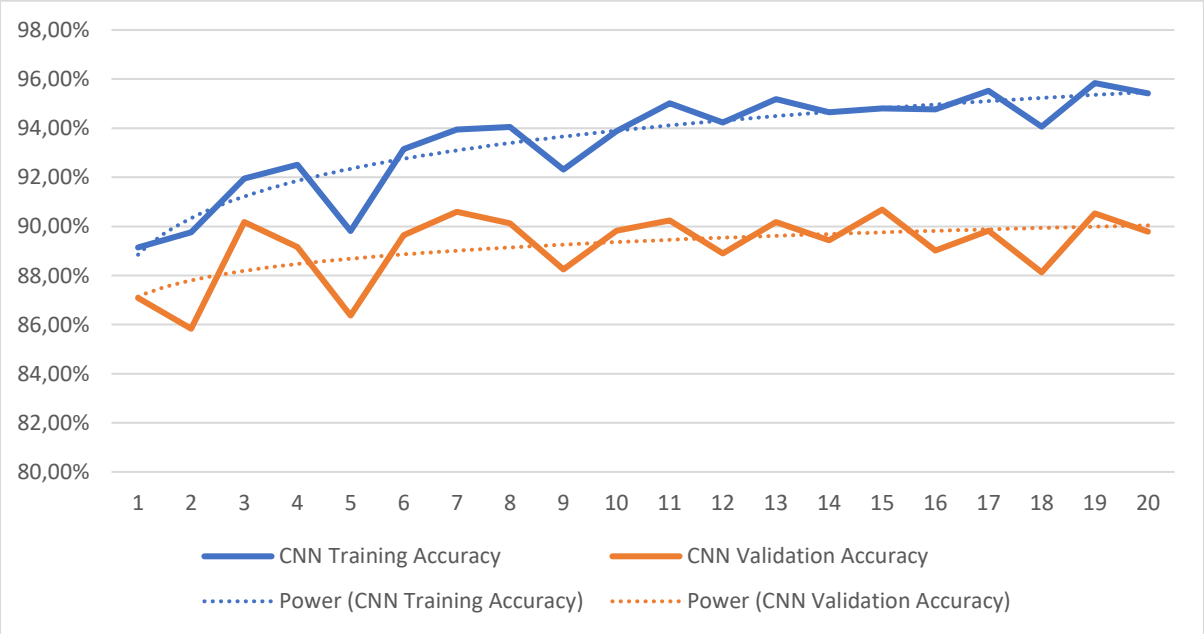


Figure 8: Calculated Accuracy of the CNN on training and validation samples between 1-20 epochs. Training accuracy increase from 87% to 95%, while validation accuracy increases from 87% to 90%. While both values increase, the curves start to flatten out and reach a plateau at around 90% (validation accuracy) and 96% (training accuracy).

The training was run for several epochs with the aim to increase the overall accuracy of the model. During the training, the training and validation accuracy kept increasing until it reached a plateau at around 95% and 90% respectively (see Figure 8).

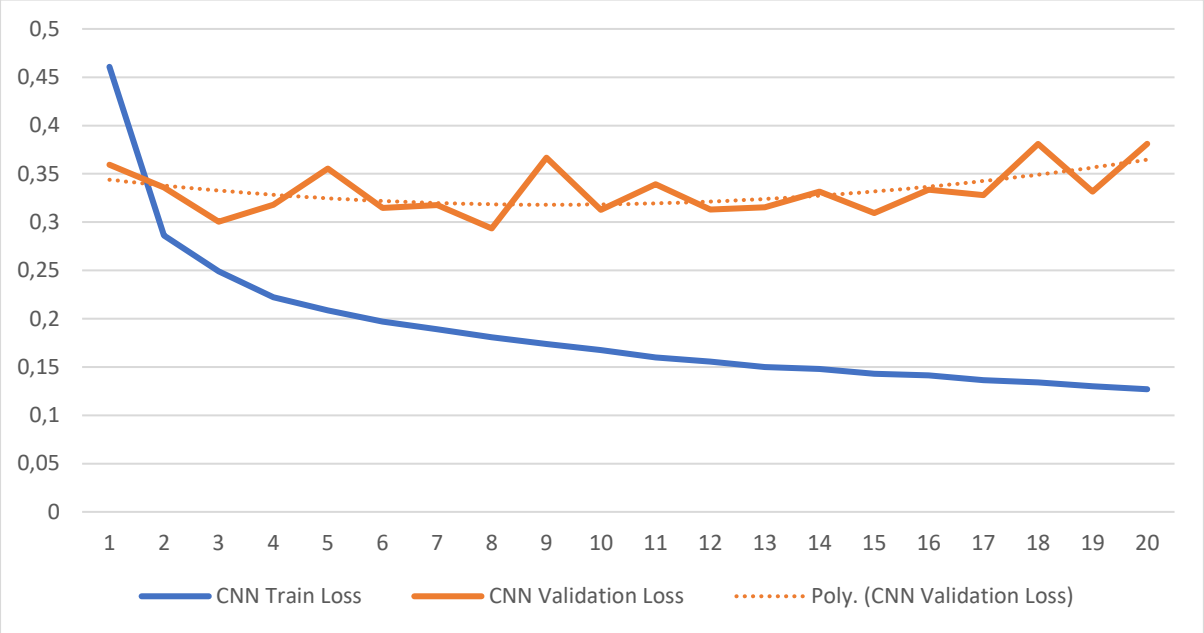


Figure 9: Performance of the CNN between 1-20 epochs. While the training loss continuously decreases from 0.46 to 0.13, validation loss stays at around 0.33.

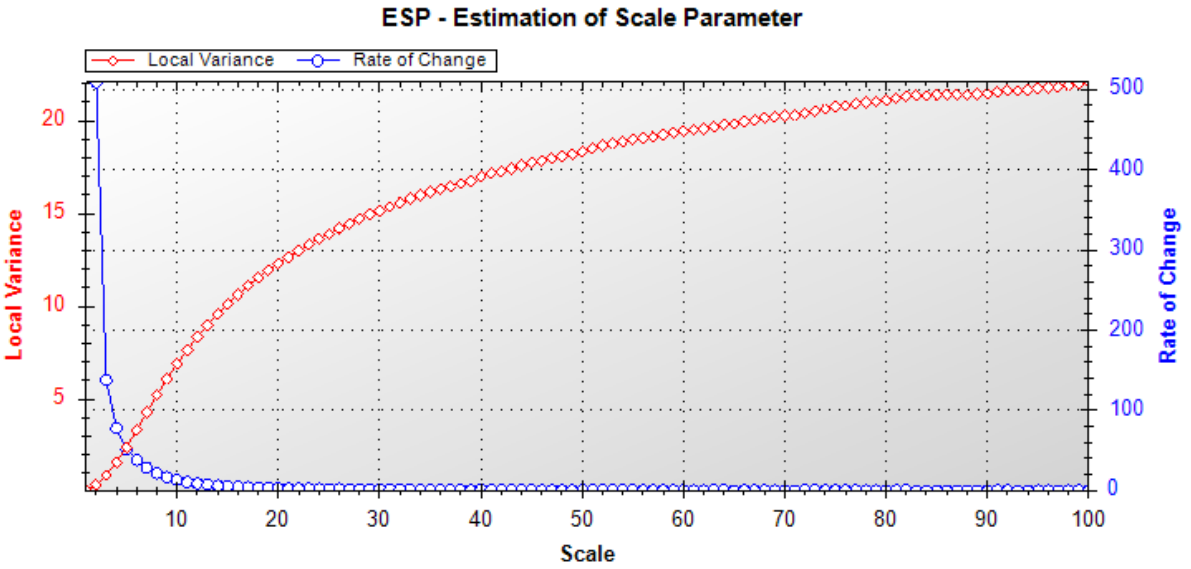
The validation loss slowly started to increase again at epoch 15 (see Figure 9), signalling increasing risk of overfitting. As overfitting leads to deteriorating generalization capabilities, training was ultimately stopped after 10 epochs, using the early stopping method described in section 2.5.2 Convolutional Neural Networks'. After concluding the training, the training statistics were exported, and the model was saved.

4.2.3.2 LULC Segmentation and Base Classification

For the classification of the LULC classes, the project was switched to the 'Salzburg-West' study area. In a preprocessing step, the brightness values of the map image were again calculated. Then, the previously trained classification model was imported and executed, creating a probability heatmap for each of the 14 training classes. To reduce potential noise and errors on the pixel level, each heatmap was also smoothened using a simple Gaussian filter with a kernel size of 15.

The OBIA classification was divided into three parts, the image segmentation, the base classification of the image objects, and the expert-based reclassification. For the segmentation of the image, the MRS algorithm was used. Advantages of this segmentation approach have been discussed in the section 2.4.2 Multiresolution Segmentation'. As the map contains image objects of different size and scale, three hierarchically structured segmentation levels were created (Level_0, Level_1, Level_2).

The Estimation of Scale Parameter (ESP) 2 tool was used to statistically determine the most suitable scales for segmentation. A shape value of 0.1 and a compactness value of 0.5 were selected for the segmentation, which maximized the weighting of colour in relation to shape. The ESP2 tool identified three optimal levels, with a scale of 433, 133, and 83, with 83 being the smallest.



Each level represented different map characteristics. The largest segmentation level, level 433, was suited to segment whole agricultural fields and other large patches. The smaller level 133 was able to represent road networks and smaller gardens and meadows, while the smallest level, level 83, was able to accurately delineate buildings and letters. As the map also contains fine lines which were not well detected by the smallest level 83, the decision was made to risk oversegmentation and use a smaller scale of 33 for the image segmentation instead. The other segmentation levels, 133 and 433 were kept unchanged.

After conducting the segmentation of the image, the map image was ready to be classified. Classification was applied to all three scales based on the probability heatmaps generated earlier, creating three different classifications. The results of the base classification were copied and exported.

4.2.3.3 Expert-based Reclassification

In order to improve the classification result of the base classification, an expert-based reclassification phase was added. All previously classified levels (0-2) were copied to a new map. As image objects occur on different levels, classification accuracy varies between individual classes. For larger map features, such as agricultural fields, classification accuracies were generally higher on larger scales. Linear features, such as roads, or features with varying size, such as bodies of water, were best represented on image object level 'Level_1', while small scale features such as buildings or cartographic letters and lines were only correctly segmented on the smallest image object level.

Table 4: Each class was assigned to one of the three image object levels where it would produce the best results.

Level_0 (scale parameter 433)	Agriculture, Forest, Garden, Meadow, Other
Level_1 (scale parameter 133)	RPath, BPath, Water
Level_2 (scale parameter 33)	BLine, RLine, NoData, RUrban, YUrban, YPath

Therefore, to improve the overall classification accuracy, a rule-based reclassification classification was conducted using all three levels. Table 4 shows the image object levels and classes used in this classification. For this the classifications on 'Level_0' from the classes 'Agriculture', 'Forest', 'Garden', 'Meadow', and 'Other' were taken over to the smallest image object level. The same was done for 'Level_1' where 'RPath', 'BPath' and 'Water' classifications were transferred to 'Level_2'. After copying the classification results of 'Level_0' and 'Level_1' to the smallest segmentation level, both were deleted. All further reclassifications were only applied to 'Level_2'.

For the refinement of the classification results additional conditions were added. The first step in the class refinement was the improvement of letters and lines. Red and black letters and lines were classified well overall, but some parts were not detected. As both classes generally have low brightness, objects with a mean brightness smaller or equal to 130 were reclassified to 'RLine' or 'BLine'. If the class 'BLine' had a mean brightness of smaller or equal to 20, then the class would be reclassified as 'NoData'. If remaining image objects of the class 'BLine' had a mean likelihood of greater or equal than 25% to be forest it was reclassified as 'Forest'. 'BLine' image objects which had a mean likelihood of greater than 10% to be forest and which bordered a forest class, were reclassified as forest. This was repeated 10 times. 'NoData' with a mean brightness greater than 20 was reclassified as one of the other LULC classes based on mean likelihood. Classes with a mean likelihood of greater or equal than 80% to be water would be classified as class 'Water'. 'Water' would be reclassified to one of the other LULC classes if the mean likelihood to be water was smaller than 10%. 'YPath' would be reclassified to 'YUrban' if it shared a relative border to 'YUrban' of greater or equal than 10%. Additionally, 'YPath' would be classified as 'YUrban' if mean likelihood was greater or equal than 20% to be 'YUrban' and the distance to 'RUrban' is smaller or equal than 300 pixels. 'RPath' was refined by reclassifying 'RPath' as 'RUrban' if the total area was smaller or equal than 3000 pixels, the distance to 'RUrban' was smaller or equal than 300 pixels, and the mean likelihood to be 'RUrban' was greater or equal than 20%. To improve the class 'Garden', image objects with a polygon compactness of smaller than 0.1 and a relative border to 'Meadow' of greater or equal than 25% would be reclassified as 'Meadow'. Similarly, 'Meadow' with a polygon compactness of smaller than 0.1 and relative border to 'Garden' of greater or equal than 25% would be reclassified as 'Garden'.

After this first reclassification round, all segments of the same class were merged using the '*merge region'* algorithm. Then, the second refinement step was conducted, improving the class 'RUrban'. Urban classes were reclassified if they were fully surrounded by classes which would be very unlikely to surround buildings, such as agricultural fields or roads. 'RUrban' with a relative border to 'Agriculture' of 100% and mean likelihood to be 'RLine' of greater or equal to 20% was reclassified to 'RLine', while 'RUrban' with a relative border to 'BPath' of 100% and a mean likelihood of greater or equal to 20% to be 'RLine' was reclassified as 'BPath'.

After applying these threshold-based reclassifications, the remaining cartographic classes were removed. As letters and lines on the scanned map were not part of LULC, they had to be removed from the classification. In the initial classification this information had been classified as 'BLine', 'RLine', and 'NoData'. To remove these classes from the map, the eCognition ruleset '*pixel-based object resizing'* was used, that allows neighbouring classes to grow pixel for pixel into the selected classes. 100 cycles were set which would assure that all areas covered by letters and lines would be eventually replaced by their neighbours. After running the ruleset, the remaining classes were merged again.

In a final step, the classification was cleaned using the '*remove objects'* ruleset. At first, 'BPath' and 'RPath' were removed if they were less than 3000 pixels in size and were 100% enclosed by the classes 'Agriculture', 'Forest', 'Garden', 'Meadow', or 'Other'. Then, 'Garden', 'Forest', and 'Meadow' were removed, if their respective pixel count was less than 3000 pixels and they were at least 70% enclosed by either 'Agriculture, 'Forest', 'Garden', 'Meadow' or 'Other'. The class 'Other' was replaced by 'Agriculture', 'Forest', 'Garden', or 'Meadow' if the pixel count was less than 5000 pixels and the relative border to one of the other classes was greater than 70%. In the end the class 'YPath' was replaced by neighbouring 'Agriculture', 'Forest', 'Garden', 'Meadow', or 'Other' if its pixel count was less than 5000 pixels and the relative border to one of these classes was greater than 70%. Then, all classes were merged once again. The reclassification result was copied to the map 'Result' and exported as a shapefile vector dataset, including the most important image object attributes, such as brightness, name, mean blue, mean, green, mean red, the mean likelihood for each class, and the number of pixels.

4.2.3.4 LULC Accuracy Assessment

After finishing the expert-based reclassification, the accuracy of the final classification result was assessed. For this task, the working environment was switched to ArcGIS Pro. The vector dataset of the classification results and the generated accuracy assessment points were imported. With the *'Compute Confusion Matrix'* geoprocessing tool, an error matrix was calculated, containing statistical information about the classification, including values for overall accuracy, producer's and user's accuracy, and Cohen's Kappa. The results were then exported.

5 Results

The following chapter presents the obtained results of the map processing workflow. Like the general image analysis workflow, the result is separated into two parts, the recognition of selected map symbols and the classification of different LULC types. All datasets, including classification results and accuracy assessments are available on the project's GitLab page[1] and on the university SharePoint[2]. The final map product for both study areas is also available in the Appendix.

5.1 Map Symbol Recognition

Map symbols are important sources of information in maps. In the Franciscean Cadastre, they provide valuable additional contextual information for various surface area classes. For example, the exact type of forest is indicated by either deciduous or coniferous tree symbols on a dark brown background, which adds further information about land cover that cannot be extracted from the map using colour information alone. Therefore, one of the research objectives in this study was to assess if map symbols could be recognized using CNNs and if they could be utilized to further improve the classification result.

For this task, six different map symbols were identified, and extracted: Coniferous (oval and spiky), deciduous (oval and round), meadows, and small shrubbery. Unfortunately, the overall results obtained by the workflow were not accurate enough to be usable for further classification tasks. While many symbols were correctly detected, several false positive identifications were also made, which reduced the accuracy of the result and its usability for further classification tasks. The identified issues were prevalent to all six map symbol classes. Meadows, for example, were often correctly detected, but the algorithm simultaneously produced several false positives as well (see Figure 10). For the other classes the overall accuracy was mixed as well, with many map symbols being correctly identified but several false positives and false negatives reducing the validity of the dataset. Figure 11 illustrates the erroneous detection of round deciduous trees near a small river. Besides detections of symbols by the CNN model, there were also issues with the identification of local maxima in the probability heatmaps which influenced the final accuracy. While most detections did not overlap each other, some areas produced multiple local maxima that were closely located next to each other, creating several points for one map symbol.

Due to the mixed accuracy and the large scope of work any subsequent refinements to the workflow were omitted and the map symbol recognition model was not applied to the second study area.

[1] https://git.sbg.ac.at/s1039004/historic-map-classification

[2] https://plusacat-my.sharepoint.com/personal/simon_meyer_stud_plus_ac_at/_layouts/15/onedrive.aspx?id=%2Fpersonal%2Fsimon%5Fmeyer%5Fstud%5Fplus%5Fac%5Fat%2FDocuments%2FHistoric%5FMap%5FClassification%2FHMC%5FProject&ga=1

Further research focused solely on the classification of LULC types, without the use of the map symbol dataset.

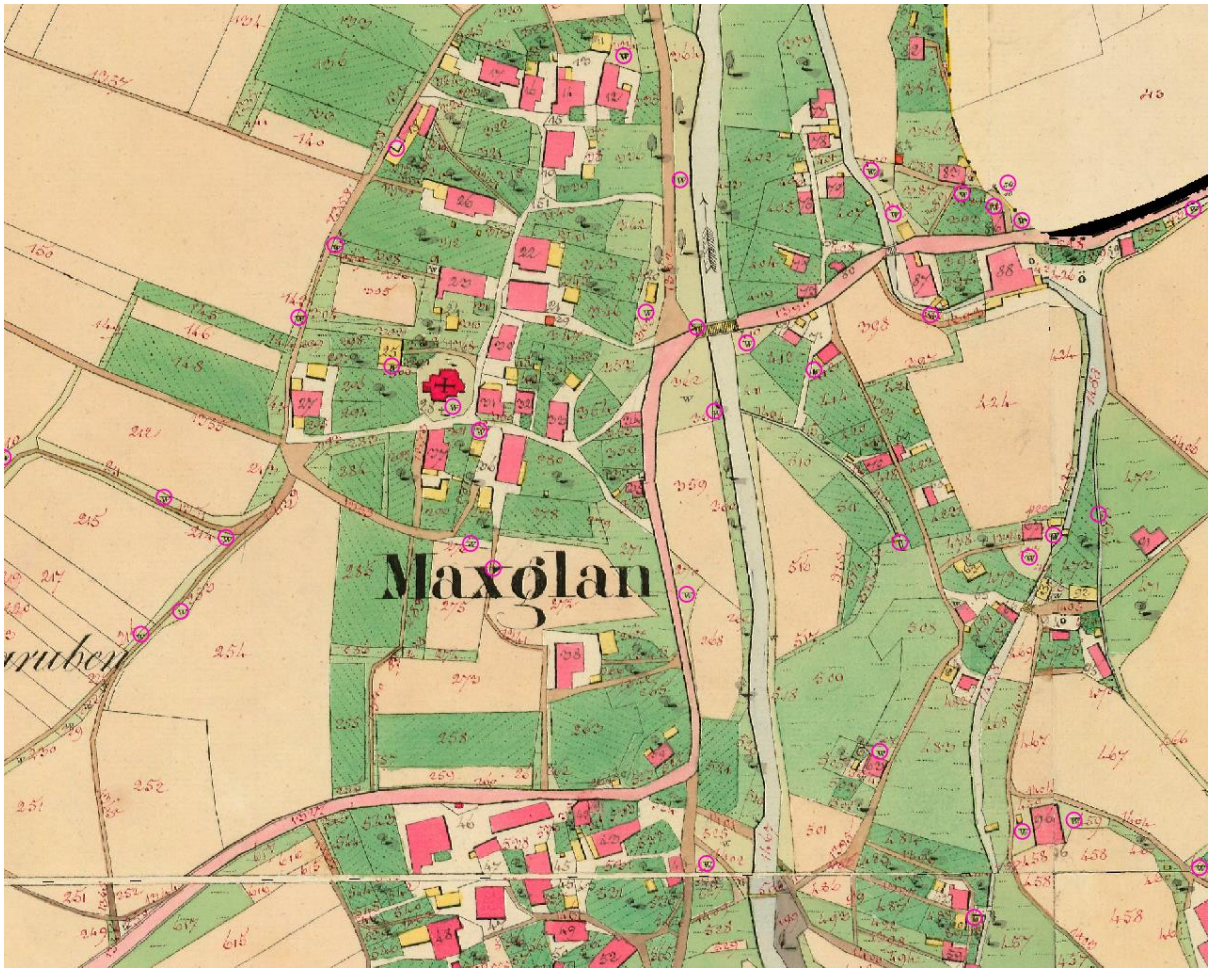


Figure 10: Map symbol recognition results in Maxglan. Meadow symbols indicated by the letter 'w', were mostly detected, but several false positives and false negatives can be identified in the image subset. False positives can be found near block letters and lines, which contain similar colour properties.

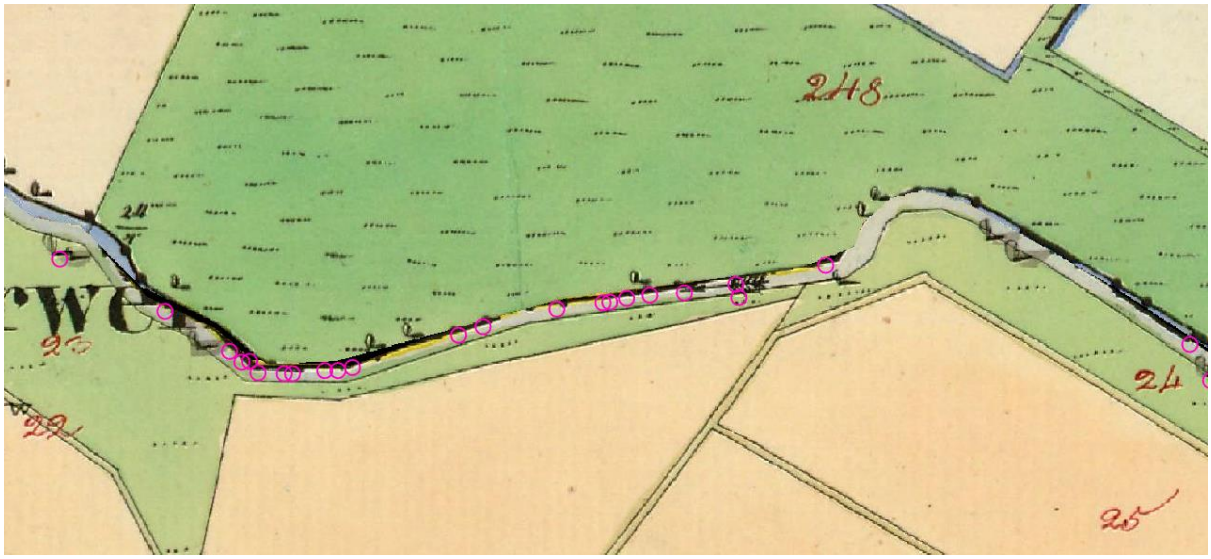


Figure 11: Falsely recognized round deciduous trees along one of the smaller rivers.

5.2 LULC Classification

The main objective of the study was the (semi-)automated extraction of geographic information from the Franciscean Cadastre. As the cadastre was not only developed for the documentation of land properties, but also for tax estimations, it includes detailed information about the usage of each plot of land. For the classification of different LULC types in the Franciscean Cadastre, 14 classes were initially selected for training, including two classes to detect black and red letters and lines in the map image, as well as one class to detect black no data regions (see Table 5).

Table 5: LULC classification classes.

LULC Classes		**Temporary Classes** (removed during classification)
▪ Agriculture ▪ Forest ▪ Garden ▪ Meadow ▪ Other ▪ Water	▪ Red urban (RUrban) ▪ Yellow urban (YUrban) ▪ Brown path (BPath) ▪ Red path (RPath) ▪ Yellow path (YPath)	▪ Black letters and lines (BLine) ▪ Red letters and lines (RLine) ▪ No data (NoData)

In addition to the three temporary classes, 11 LULC classes were chosen to represent the various surface types documented in the cadastre, as well as different road types and buildings (see Figure 7). Classifications were conducted for both study areas. The first study area around Salzburg was used to test and improve the classification workflow, while the second study area between Rauris and Taxenbach was mainly used to assess the transferability of the approach to different sections of the cadastre, especially to areas with a very low quality of the physical map source.

Overall, the results of the LULC classification are satisfactory for the Salzburg area, but significantly less accurate in areas with degraded map quality, especially in the second study area. In the Salzburg study area, most classes were successfully delineated and classified, and only minor, small scale misclassifications were present in areas with blurry boundaries. A main classification problem was the differentiation between the class ‘Water’ and ‘Other’, as both classes lack a distinct colourization. The class ‘Water’ is frequently not presented with a contiguous colourization, but by a fine blue colourization of the class borders instead. This results in misclassifications by the CNN in the central areas of the class which are then wrongly assigned as ‘Other’, a class designed to represent non-classified areas on the map. Classification results for both study areas are shown in Figure 12 and Figure 13, which visualizes the problem along the Salzach river and at severely degraded map sheets of the second study area.

Figure 12: Final LULC classification for the Salzburg area. The area consists of two map images ('Salzburg-West' and 'Salzburg-East') that were processed and validated separately and then merged in GIS. The 'Salzburg-East' area was used for the training of the classification model, while the 'Salzburg-West' area was used in the development of the classification workflow and the final accuracy assessment.

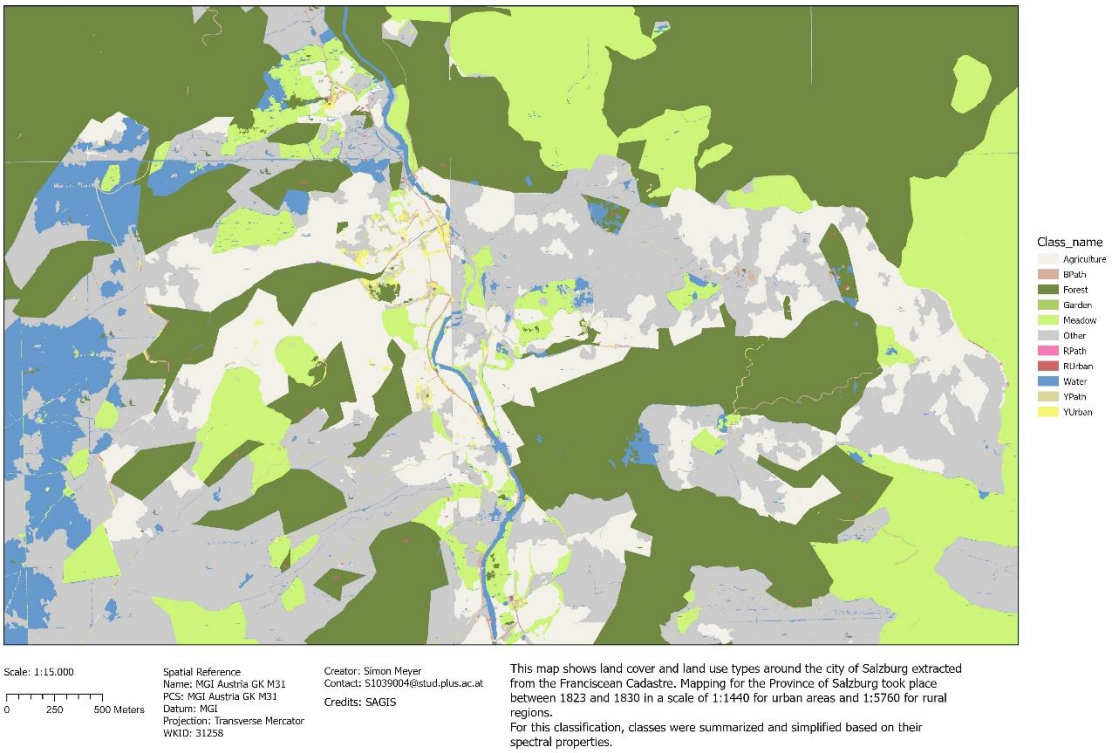


Figure 13: Final LULC classification for the second study area between Rauris and Taxenbach. The overall classification accuracy is considerably lower than for the Salzburg study area, due to significant decolourization and paper yellowing of the original map source. Although the accuracies are low, this map section presents the worst-case scenario for the classification model and most other regions will achieve considerably higher classification results.

5.3 LULC Classification Accuracy Assessment

The perceived high overall accuracy of the LULC classification for the Salzburg area was supported by a statistical accuracy assessment. A 90.5% overall accuracy was achieved on the 'Salzburg-West' classification area with a Kappa index of 86% (see Table 6), based the total testing sample count of 526 samples.

Table 6: Confusion Matrix of the 'Salzburg-West' study area.

Class Name	Water	Meadow	Forest	Garden	Agriculture	Other	RPath	BPath	YUrban	RUrban	YPath	Total	U_Accuracy	Kappa
Water	10	0	0	0	0	0	0	0	0	0	0	10	1	0
Meadow	0	106	0	0	2	2	0	0	0	0	0	110	0,96363636	0
Forest	0	0	28	0	0	0	0	0	0	0	0	28	1	0
Garden	0	0	1	16	0	0	0	0	0	0	0	17	0,94117647	0
Agriculture	0	2	0	0	263	0	0	0	0	0	0	265	0,99245283	0
Other	19	1	0	0	1	19	0	1	0	0	0	41	0,46341463	0
RPath	0	3	0	0	1	0	5	0	0	1	0	10	0,5	0
BPath	0	2	0	0	1	0	0	12	0	0	0	15	0,8	0
YUrban	0	0	0	0	0	0	2	0	8	0	0	10	0,8	0
RUrban	0	0	0	0	0	0	2	0	0	8	0	10	0,8	0
YPath	0	9	0	0	0	0	0	0	0	0	1	10	0,1	0
Total	29	123	29	16	268	21	9	13	8	9	1	526	0	0
P_Accuracy	0,34482759	0,86178862	0,96551724	1	0,98134328	0,9047619	0,55555556	0,92307692	1	0,88888889	1	0	0,90494297	0
Kappa	0	0	0	0	0	0	0	0	0	0	0	0	0	0,8611595

While the overall classification accuracy was high, the accuracy per class varied considerably. Most notably 'Water' had a low accuracy, only reaching a producer's accuracy of ~35%. Almost 60% of generated 'Water' samples had mistakenly been classified as 'Other', supporting the observation made earlier. 'RPath' reached a producer's accuracy of ~56%. The rest of the classes reached a producer's accuracy between 86% for the class 'Meadow' and 100% for the classes 'Garden', 'YUrban', and 'YPath'. 100% accuracies were mainly achieved in classes with low sample counts. The class 'Garden' with 16 samples, reached 100% producer's accuracy. The largest land cover class 'Agriculture' with a total sample count of 268 reached a producer's accuracy of 98%, while the second largest land cover class 'Meadow' reached 86% producer's accuracy.

User's accuracy also varied. The lowest user's accuracy was achieved by the class 'YPath', which is only present in a small section of the map. Because the class has similar colour properties as the class 'Meadow' it was often falsely classified as such, resulting in a user's accuracy of only 10% and multiple false positives. Similarly, the class 'Other' only reached a user's accuracy of 46%, often being misclassified as 'Water'. The user's accuracy of 'RPath' was also low with 50%, often being mistakenly classified as 'Meadow', 'Agriculture', or 'RUrban'. The other classes reached significantly higher accuracies, with 'Water', 'Meadow', 'Forest', and 'Agriculture' each achieving more than 95% user's accuracy. The class 'Water' is again notable, because even though its producer's accuracy was low

overall, all detected samples were correctly classified, leading to no false positives and a 100% user's accuracy.

Although the Salzburg study area showed a high overall accuracy in the LULC classification, the result could not be repeated in the second study area between Rauris and Taxenbach (see Table 7). With a total accuracy of only 31% and a Kappa index of 17%, no satisfactory result could be obtained.

Table 7: Confusion Matrix of the study area between Rauris and Taxenbach.

Class Name	Water	Meadow	Forest	Garden	Agriculture	Other	RPath	BPath	YUrban	RUrban	YPath	Total	U_Accuracy	Kappa
Water	0	27	2	0	0	0	0	1	0	0	0	30	0	0
Meadow	0	85	0	1	0	0	0	0	0	0	0	86	0,98837209	0
Forest	6	96	73	0	0	0	0	2	0	0	0	177	0,41242938	0
Garden	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Agriculture	0	70	2	0	0	0	1	1	1	0	0	75	0	0
Other	9	107	2	1	0	2	0	2	2	0	1	126	0,01587302	0
RPath	1	5	0	0	0	0	4	0	0	0	0	10	0,4	0
BPath	0	7	0	0	0	0	0	2	1	0	0	10	0,2	0
YUrban	0	3	0	1	0	0	0	1	5	0	0	10	0,5	0
RUrban	2	3	5	0	0	0	0	0	0	0	0	10	0	0
YPath	0	9	0	1	0	0	0	0	0	0	0	10	0	0
Total	18	412	84	4	0	2	5	9	9	0	1	544	0	0
P_Accuracy	0	0,20631068	0,86904762	0	0	1	0,8	0,22222222	0,55555556	0	0	0	0,31433824	0
Kappa	0	0	0	0	0	0	0	0	0	0	0	0	0	0,17044693

However, the confusion matrix reveals that classification inaccuracies are not distributed equally between all detected classes. As in the Salzburg area, the individual class accuracies varied, with producer's accuracies ranging from ~21% to 100%. Low producer's accuracies were obtained for the class 'Meadow' with ~21%, 'BPath' with 22%, and 'YUrban' where only 55% were classified correctly. 'RPath' and 'Forest' achieved higher producer's accuracies with 80% and 87% respectively. The remaining classes either achieved 0% or 100% accuracy, but with low or no sample counts in this region.

User's accuracy varied as well, with five classes reaching 0% accuracy, including 'Water', 'Garden', 'Agriculture', 'RUrban', and 'YPath'. 'RPath', 'BPath', and 'YUrban' achieved user accuracies ranging from 20% to 50%. The highest user's accuracy was obtained with the class 'Meadow' with 99%, indicating, that if the class was detected it was mostly classified correctly. The other frequently classified classes 'Forest' and 'Other' reached user's accuracies of 42% and 1.6%. The low user's accuracy for the class 'Other' indicates that the algorithm falsely classified many areas that were meadow as 'Other'.

6 Discussion

Although the overall accuracies for the LULC classification were high in the main study areas, significantly lower accuracies were obtained when the classification workflow was applied to the low quality map sections in the High Tauern region. The following chapter will interpret the visual and statistical results obtained and discuss strengths and limitations of the studies' methodology.

6.1 Map Symbol Recognition

The findings outlined in 5.1 Map Symbol Recognition', demonstrate the (theoretical) feasibility of utilising CNNs to detect and extract map symbols from the Franciscean Cadastre. While the overall accuracy was comparatively low due to the presence of several false positives, the CNN was able to differentiate the individual classes. Complex symbols, such as the letter 'w' representing meadows, were often successfully detected. Also, larger objects such as coniferous trees were often successfully identified, but a high number of false positives persisted which reduced the model's overall validity.

As the number of unique training samples for the training of the CNN model was low (see Table 1), the model's learning capabilities were restricted. While data augmentation was able to increase the number of samples to a level where the model was able to identify many of the symbols, the recognition capabilities most likely remained insufficient for the detection of most complex symbols. Potential model improvements were also confined by the limited augmentation options available for map symbols. Mirroring and rotation of the samples was largely not an option because the orientation of map symbols is crucial for class differentiation. The only augmentation options left for the sample patch generation was restricted to slight rotations which was not sufficient.

Future research should therefore start by increasing the number of unique training samples and a larger training area with more map sheets of the cadastre to capture a greater variety of symbols. Besides the use of a larger training dataset, increasing the neural networks depth and number of feature maps could also help to improve the overall accuracies. In addition to remodelling the CNN training process, alternative classification methods, such as template matching, could potentially yield better results with low sample counts. This option should be explored, before investing additional time into the training of a new CNN classification model. In the past, template matching algorithms and have been explored for the extraction of text and symbols from historic maps (Budig and van Dijk, 2015, Xia et al., 2022) and implementations are available to eCognition (Trimble, 2024a).

6.2 Map Segmentation

Crucial prerequisite for successful OBIA, is the accurate segmentation of the image into meaningful image objects, that represent the features in the image. The segmentation results by the MRS did overall represent the class boundaries of most of the image objects found in the original map source.

Results were generally very precise in areas with large spectral contrast, but if the colour contrast was low, the multiresolution segmentation algorithm sometimes created segments that included different objects, such as trees symbols or lines, resulting in inaccurate classifications where objects extended into neighbouring classes. While these issues persist in the entire study area, they generally only influence small fragments of the map, and they largely did not impact overall the classification results.

Minor segmentation issues were located the roads where parts of the roads were seemingly growing into the neighbouring fields, as well as some tree symbols near the roads, which were mistakenly included in some of the road segments (both visible in Figure 14). This created some imprecise borders for sections of the road network, which could theoretically be smoothened with additional morphology tools, but were kept in the final classification, as morphology tools will fluence other map areas as well.

Figure 14: Comparison of classification results near Aiglhof, Salzburg.

Low colour differences and subsequent segmentation errors were also identified between gardens, meadows, and brown paths, where the segmentation led to suboptimal boundary delineations. In Figure 14 some misclassifications can be found where the multiresolution segmentation algorithm created segments that do not represent actual map features. These errors were then transferred to the final classification result. In general, these issues are known issues of the MRS algorithm and have been described by Ma et al. (2017) before.

Nevertheless, most classes were segmented correctly. Figure 15 shows a detailed section of the historic city centre of Salzburg, where letters overlap and obstruct building footprints. Even though the scene contains many complex boundaries, the final classification was able to eliminate most of the cartographic lines and letters, which indicates that the chosen segmentation level was sufficient to

segment most of the features. Finally, it must be noted, that while some minor segmentation errors persisted, the resolution of the classification result is in general fine enough that negative impacts of the segmentation on the final classification can at the most part be neglected.

Figure 15: Comparison of the classification results near Staatsbrücke, Salzburg.

6.3 LULC Classification Accuracy

As outlined in section 4.2.3.4 LULC Accuracy Assessment', the overall classification accuracy achieved for the main study area around Salzburg was high, but individual class accuracies varied greatly. The two main reasons for inaccuracies in the classification result are either spectrally similar classes, that were not correctly classified by the CNN and subsequent reclassifications, or too complex classes, that require specifically optimized classification models and additional object-based classification algorithms. Several reasons for spectral similarity between classes have been identified:

One reason is, that the map source, like many historical maps, has varying grades of quality. Some map sheets are significantly degraded and contain regions with low colour differences. In these regions colours either faded or turned yellow or brownish due to the aging paper. This issue is transferred to the digital dataset in the map scanning and digitization process. The scanning can also worsen the problem and introduce additional artifacts, such as halos along edges, over sharpening and additional discolouration (see section 1.3 Current Challenges in (Semi-)Automatic Map Processing

1.3 Current Challenges in (Semi-)Automatic Map Processing'). All these errors and issues in map quality can influence the results of both the image segmentation as well as the image classification process. While the CNN can potentially learn some of these maps colour characteristics and reduce their impact on the classification, some errors will find their way into the final classification result. The following sections will discuss the LULC classification results, as well as the removal of cartographic symbols from the final result.

6.3.1 Class Accuracy

As described previously, the results of the LULC classification in the 'Salzburg-West' study area had for the most part in high accuracies. Even though the overall accuracy was high, significant differences between individual class accuracies could be observed.

For the buildings classes, the segmentation and classification did provide good results, due to their high colour contrast compared to most other classes and the black outlines present in the original map source. After the reclassification phase, there were surprisingly no large issues in separating red lines from red buildings. Minor problems were identified when removing black lines and letters from the map, which did sometimes create uneven borders which slightly deviate from the original class borders. Utility buildings (coloured in yellow) were also classified well.

Another challenge in map analysis process had been the detection of agriculture, which was important as this class covers large areas of the Salzburg study area. Most agricultural fields were classified well, but during the training and the creation of the probability heatmaps some minor issues had been observed, where areas neighbouring the class 'Other' had been incorrectly detected as such, probably due to their colour similarity. This issue in the classification model was largely solved during the reclassification phase, where agriculture was simply classified on a larger image object level. A similar classification issue was observed in areas next to meadows, especially when the thin meadow strips along the agricultural fields were not well coloured. No simple rule for reclassification was found, and some misclassifications remain. However, these only make up a small area of the study area.

Meadows, like agricultural fields, were overall well classified. Minor problems were observed where meadows border gardens. While most areas were correctly identified as garden or meadow, minor misclassifications due to low colour differences were observed.

Gardens were predominantly classified correctly. Small scale misclassifications were observed, when located next to meadows. As both classes have similar colours, an improved classification algorithm might provide better results. In addition, larger sample patches might help the CNN to gather more spatial context and thereby detect the pattern of gardens, improving the distinction between both classes.

Forests were classified very well as they are with their dark brown colour distinct from most other classes. Minor issues were found in small patches around the ‘Mirabellgarten’, where the small segment sizes in combination with the presence of several classes resulted in low probabilities and thereby low confidence classifications with higher error rates.

The class ‘Other’ was mostly correctly classified, apart from the issues with the differentiation between the ‘Water’ and ‘Agriculture’ classes (see Figure 16). The biggest issue here, was to prevent the classification of areas which are clearly not water as ‘Water’. This was achieved, but with the consequence of a very low accuracy for the class ‘Water’. Water was mostly derived well in areas that were already by colour clearly marked as ‘Water’, such as small rivers, creeks, and most borders along larger bodies of water. Lower overall accuracies of this class were accepted if it resulted in fewer false positives and higher accuracies of other classes. The class ‘Water’ was thereby the only class, that requires further manual input after running the entire classification, which is indicated by the low final producer’s accuracy of only 35%.

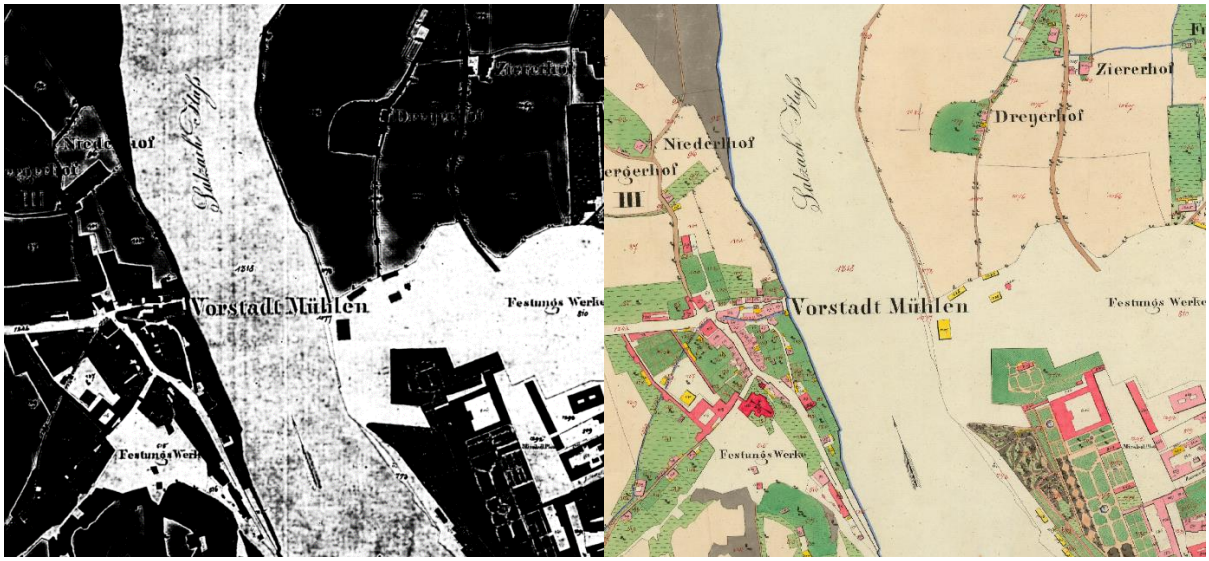


Figure 16: Probability heatmap of the class 'Other' compared to the original map source. In areas which border agriculture and along black lines, the classification model falsely identifies higher likelihoods of the class ‘Other’, which can influence the final classification. The heatmap also visualizes the model’s confusion between uncoloured bodies of water and ‘Other’.

Besides buildings and vegetation, roads make up a considerable part of the overall surface area. The class ‘BPath’ represents simple paths and was mostly well detected and classified. Some issues did persist and could not be completely solved, such as small issues with the correct segmentation from black symbols and lines. 'RPath' was also mostly correctly classified, but like other road classes, its accuracy was affected by incorrect segmentations. Its spectral similarity to red lines and red buildings sometimes led to an incorrect classification and it is important to note that not all issues were able to be solved by additional conditions during the reclassification phase.

In difference to the other two road classes, 'YPath' was only found in one small area in the ‘Salzburg-West’ study area. Although the class was correctly detected and classified, false positives were more common due to its spectral similarity to meadows in other areas. Additional expert-based rules were unable to completely solve the issue and might require adjustments to the classification model.

Besides the colour similarities between several classes, the image interpretation was also a source of error. The Franciscean Cadastre includes several LULC classes that are not solely characterized by one colour or shape, but instead are indicated by multiple colours, textures, and complex shapes. For example, quarries are classified by a blueish-grey stone texture, clay pits are represented by oval brown-black clay textures, and peat cuttings are symbolized by rectangular brown bricks. This is challenging to model, as it traditionally requires high-level reasoning. Due to time constraints, this study did not aim to classify all complex classes of the Franciscean Cadastre, which may be a task for future studies.

However, significant problems were encountered when classifying and separating the class 'Water' from 'Other'. While water is generally coloured blue, larger areas are like the class 'Other' uncoloured. For larger bodies of water, boundaries are only implied by a thin blue colour strip at selected sections of the class border. This creates issues with the correct training and recognition capabilities of the classification model. In this study therefore, training samples were only taken in areas with visible blue colouring, to avoid issues with separating both classes. The final classification was intended to be conducted using either thresholds and additional geographical conditions in the object-based reclassification phase or a type of growing algorithm, both of which weren’t successful because of several reasons. The reclassification approaches were unsuccessful due to several reasons. Both classes border each other at the city centre of Salzburg, where fortifications classified as ‘Other’ directly border the Salzach river. The classes are separated by a thin black line, which cannot be continuously delineated by current classification methods. Region-growing algorithms which aim at extending the ‘Water’ class into neighbouring classes, however, were growing into the fortification area, which made accurate classifications impossible. Further attempts did not yield good results, resulting in the current low producer's accuracy for the 'Water' class and the necessity of manual reclassifications.

6.3.2 Removal of Letters and Lines

The removal of all boundaries and text overlays from the classification result is a crucial step in the digitization and vectorization of LULC information from historic maps. Within the map processing workflow, the removal was part of the reclassification phase after the map had been segmented and classified. A pixel-based object resizing algorithm was used to remove letters, lines, map symbols, and other non-land-cover areas from the final classification. The algorithm successfully removed the black

outlines around buildings and parcels while retaining the overall shape of the building footprints and other smaller objects. Larger objects, such as place and town names, and no data regions, including small gaps between the old map sheets, were also removed. The cleanup of letters and lines was done in both urban and rural areas.

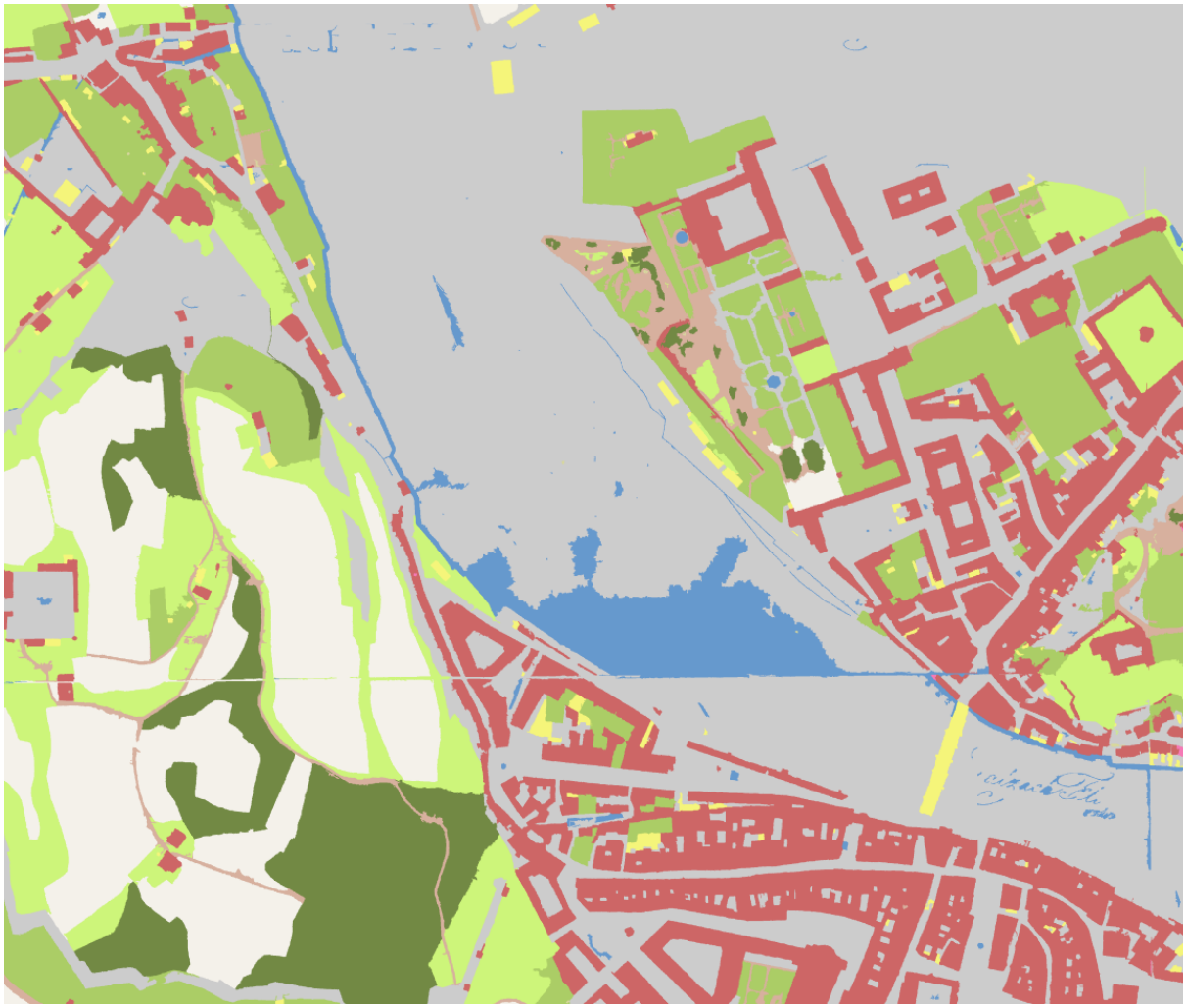


Figure 17: Misclassifications at the Salzach river. Some outlines around letters remain, as they were falsely classified as water and subsequently missed in the reclassification phase, which aimed at removing letters, lines, and no data areas from the classification. A horizontal line is also visible, which was likely caused by the mosaicking of the map sheets after the scanning of the original map source.

Smaller objects, including various map symbols coloured in black and parcels numbers coloured in red, were successfully removed as well. The parcel numbers were initially a topic of concern, as they had a similar colour as red buildings but as they were mostly correctly detected and classified by the CNN it no longer proved to be a problem during the removal. Although most of the objects were successfully removed, some letters and lines that were previously misclassified remained. This issue was observed in some letters in the Salzach river (see Figure 17) that had been misclassified as water. However, since they were exceptions and further improvements were difficult and time-consuming, no further corrections were made.

6.4 Strengths and Weaknesses of the Classification Approach

Historic maps present several unique challenges to classification algorithms. The decision to use OBIA combined with CNN classification to extract geographic information from the digitized Franciscean Cadastre had several advantages and disadvantages compared to alternative approaches, which will be explained below.

The evaluation of the classification approach must be carried out in a slightly different way compared to remote sensing image analysis, as there are several fundamental differences. Firstly, when analysing historical maps, there is usually no other data to support the classification. When analysing remotely sensed imagery, there is often supplementary raster or vector data available that can be used to improve the classification results. In the analysis of historic maps, the original data source is generally the only source available. The general condition of the physical map and the digitization and preprocessing process therefore have much influence on the final classification result. If the quality of the input image is poor, then even the sophisticated classification algorithms will not be able prevent false classifications. The second difference to remote sensing image analysis is located in the heterogeneity of the different historic map products. Different colourizations, map symbols, hand-written text, and the general image complexity create challenges to any current classification approach. While there are several efforts to develop fully automatic map processing, most current approaches are semi-automatic and map-specific, including the approach taken in this study.

The object-based analysis had several advantages, compared to traditional pixel-based approaches as outlined in chapter 2 Theoretical Principles of Map Processing'. The integration of expert-based rules into the classification process made it possible to define additional conditions for the classification of selected LULC types, such as the geometric shape of classes or neighbourhood relationships. By using the MRS algorithm for the segmentation of the image, small and large map features could be classified at the different scales and thereby ensure that both small map symbols and large fields could be correctly classified. In combination this allowed for high classification accuracies.

The use of a CNN model for the classification, instead of more traditional machine-learning classifiers, had several advantages. As historic map processing is highly dependent on the quality of a single data source, the digitized map, it can only use limited map characteristics in the classification. But the most common property, colour, is often unusable due to various factors such as paper aging or stains, which will mislead classifiers that only look at colour. CNNs have the advantage of being able to recognize complex patterns and based on colour and shape properties, which makes them favourable for historic map processing. Downsides of a CNN based classification approach are, time consuming creation of training samples, the parametrization, and the extended knowledge necessary (Gu et al., 2018).

In retrospect, many of the map processing specific issues listed in section 1.3 Current Challenges in (Semi-)Automatic Map Processing' were also experienced during this study. Especially, the low map quality did pose a challenge in the classification of map features. The preprocessing necessary for the training of the CNN was also considerable, with much time spent on the improvement of the model's architecture and parameters.

6.5 Transferability

The transferability of the workflow is one of the most important criteria for (semi-)automated image processing approaches. Transferability ensures the automation of tasks on other areas outside the training region. It therefore allows for scaling and increased efficiency, which are both primary objectives in the development of automated rulesets.

To test the transferability of the approach to other (degraded) map sections of the Franciscean Cadastre and assess the robustness of the image analysis approach, the workflow was applied to the second study area in the High Tauern between Rauris and Taxenbach. The Results, as expected, were less accurate, especially in heavily degraded and discoloured areas. As outlined in section 5.3 LULC Classification Accuracy Assessment', the overall accuracy per class varied. While forests were mostly classified correctly, accuracies for other classes fell significantly. Especially, grassland which makes up a huge part of the study area, was often mistakenly classified as 'Other' or 'Agriculture' in areas where the green colour had faded or yellowing of the paper had occurred.

Another potential source of error in regions with very different colour properties might be caused by the hard thresholds introduced in the reclassification section. While the thresholds did help in the 'Salzburg-West' study area to improve class accuracy, they might reduce the accuracy in other regions. The introduction of fixed thresholds should therefore be carefully implemented and potentially revised in follow-up studies.

Finally, the results in both study areas show that the current classification model and workflow are not yet good enough to overcome all challenges present in the Franciscean Cadastre. Although these results are unfortunate, it is important to note that some of the heavily degraded map sections are difficult to classify even with human interpretation. And while the transferability to heavily degraded map sections is limited; the image analysis workflow was robust enough to achieve high accuracies around Salzburg.

7 Conclusion & Outlook

Much research has been conducted in recent years to automate the information extraction process from historic map sources, driven by advances in image analysis and the continuing demand for historic geodata. In recent years, breakthroughs in deep learning techniques have demonstrated their potential for object detection and image classification tasks. While some studies already explored the potential of CNN classifiers for the classification of historic maps, their application to scanned maps has so far been limited.

This study aimed at developing a workflow for the semiautomatic extraction of geographic information from the Franciscean Cadastre using OBIA and CNNs to classify and vectorize historic maps. Research goals were twofold. The study should assess the capabilities of a combined OBIA and CNN classification approach to extract the most important features of the cadastre, especially in areas with bad image quality. In addition, the workflow was tested on a different map section to assess its transferability. The study results should provide the foundation for future extended vectorization works that would enable the quantitative analysis of the historic landscape of Salzburg and its transformation over past 200 years.

The image analysis was divided into two parts: the detection and recognition of selected map symbols which aimed to extract point data, and the segmentation and classification of important surface information. Symbol extraction results were mixed, with many types identified correctly, but the classification model was unable to achieve sufficient accuracy to use the results in the subsequent LULC classification. The LULC classification workflow was able to correctly segment and classify the most important LULC types in the ‘Salzburg-West’ study area.

The three-level multiresolution segmentation approach used for the LULC classification was successful in delineating smaller features, such as letters and lines, as well as larger objects, including agricultural fields and bodies of water. The ESP2 tool that was used to identify best-fitting scale parameters reduced the overall time necessary for image segmentation. While the segmentation was overwhelmingly successful, some minor issues with the region-based segmentation technique persisted, such as lower accuracy in low contrast regions and the unintended merging of very fine features with neighbouring classes that had similar colour properties.

Some issues with the LULC classification accuracy were observed in areas that had not been colourized properly in the original map source. These issues persisted and could not be solved entirely neither by the CNN classifier nor by the expert-based classification. When applied to the study area in the High Tauern, the LULC classification accuracy per class dropped substantially, which indicates general overfitting of the classification model and limits the transferability of the model to less degraded map

sheets. To counter issues of the supervised classification model and increase the accuracy of some LULC classes, expert-based rules were introduced. They were able to correctly remove most letters, lines, and text from the final classification result and were also able to remove some minor misclassifications. On the downside they limited the transferability of the approach as they introduced hard thresholds into the classification. Despite the named shortcomings, the OBIA approach proved to be valuable as it provided a flexible framework for the refinement of the initial classification results after the CNN training had been finished.

The CNN model was able to detect the most important LULC classes confirming the capabilities of CNN classifiers for the analysis of historic maps. It should be noted that both the LULC classification model and the map symbol detection model would likely benefit from a larger training dataset. Enriching both training datasets with more samples and adding further data augmentation methods would likely further enhance the accuracy of the models and increase its generalization capabilites in other cadastral regions. This underscores the importance of an expansive and diverse training set for comprehensive model performance.

The study's results suggest that despite recent advancements in image analysis, the digitization of historic maps remains a challenge, especially when original map sheets are of poor quality. Some degree of human intervention in the classification process is still necessary and cannot be entirely avoided, requiring further research into automated image processing. Future research on historic map analysis should therefore focus on two objectives: Firstly, increasing the number and variety of training samples by expanding the total size of the training areas. This will likely improve overall classification results and and increase model robustness in poor map conditions. Secondlys the overall accuracies of the classification results around Salzburg were already suitable for quanitative research, provided some manual adjustments are made to the classification of water bodies, classification work should also be upscaled to cover larger areas of the cadastre. By building upon the groundwork laid by this study, a large-scale approach can be undertaken to further vectorize the Franciscean Cadastre and to make important historic geographic information available for GIS analysis.

References

BAATZ, M. & SCHÄPE, A. Multiresolution Segmentation : an optimization approach for high quality multi-scale image segmentation. 2000.

BAUER, M. 2017. Die Schätzungsoperate des Franziszeischen Katasters als agrarhistorische Quelle. *Jahrbuch für Geschichte des ländlichen Raumes,* 14**,** 216-250.

BICIK, I., GABROVEC, M. & KUPKOVA, L. 2019. LONG-TERM LAND-USE CHANGES: A COMPARISON BETWEEN CZECHIA AND SLOVENIA. *Acta Geographica Slovenica-Geografski Zbornik,* 59**,** 91-105.

BLASCHKE, T. 2010. Object based image analysis for remote sensing. *ISPRS JOURNAL OF PHOTOGRAMMETRY AND REMOTE SENSING,* 65**,** 2-16.

BLASCHKE, T., HAY, G. J., KELLY, M., LANG, S., HOFMANN, P., ADDINK, E., FEITOSA, R. Q., VAN DER MEER, F., VAN DER WERFF, H., VAN COILLIE, F. & TIEDE, D. 2014. Geographic Object-Based Image Analysis - Towards a new paradigm. *ISPRS JOURNAL OF PHOTOGRAMMETRY AND REMOTE SENSING,* 87**,** 180-191.

BLASCHKE, T., LANG, S., LORUP, E., STROBL, J. & ZEIL, P. 2000. Object-Oriented Image Processing in an Integrated GIS/Remote Sensing Environment and Perspectives for Environmental Applications. 2.

BLASCHKE, T. & STROBL, J. 2001. What's wrong with pixels? Some recent developments interfacing remote sensing and GIS. *Geo-Informations-Systeme,* 14**,** 12-17.

BUDIG, B. & VAN DIJK, T. C. 2015. Active Learning for Classifying Template Matches in Historical Maps. *DISCOVERY SCIENCE, DS 2015.*

CHEN, Y., CARLINET, E., CHAZALON, J., MALLET, C., DUMÉNIEU, B. & PERRET, J. Combining Deep Learning and Mathematical Morphology for Historical Map Segmentation. *In:* LINDBLAD, J., MALMBERG, F. & SLADOJE, N., eds. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2021. Springer Science and Business Media Deutschland GmbH, 79-92.

CHIANG, Y.-Y., DUAN, W., LEYK, S., UHL, J. & KNOBLOCK, C. 2020a. Historical Map Applications and Processing Technologies.

CHIANG, Y.-Y., DUAN, W., LEYK, S., UHL, J. & KNOBLOCK, C. 2020b. Training Deep Learning Models for Geographic Feature Recognition from Historical Maps.

CHIANG, Y.-Y., LEYK, S. & KNOBLOCK, C. 2014. A Survey of Digital Map Processing Techniques. *ACM Computing Surveys,* 47.

CONGALTON, R. G. 1991. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment,* 37**,** 35-46.

DOLEJŠ, M. & FOREJT, M. 2019. Franziscean Cadastre in Landscape Structure Research: A Systematic Review. *Quaestiones Geographicae,* 38**,** 131-144.

DRĂGUŢ, L., CSILLIK, O., EISANK, C. & TIEDE, D. 2014. Automated parameterisation for multi-scale image segmentation on multiple layers. *ISPRS Journal of Photogrammetry and Remote Sensing,* 88**,** 119-127.

DRĂGUŢ, L., TIEDE, D. & LEVICK, S. R. 2010. ESP: a tool to estimate scale parameter for multiresolution image segmentation of remotely sensed data. *International Journal of Geographical Information Science,* 24**,** 872.

FOSKI, M. & LAMOVSEK, A. Z. 2019. MONITORING LAND-USE CHANGE USING SELECTED INDICES. *Acta Geographica Slovenica-Geografski Zbornik,* 59**,** 161-175.

GABROVEC, M., BICIK, I. & KOMAC, B. 2019. LAND REGISTERS AS A SOURCE OF STUDYING LONG-TERM LAND-USE CHANGES. *Acta Geographica Slovenica-Geografski Zbornik,* 59**,** 83-89.

GABROVEC, M. & KUMER, P. 2019. LAND-USE CHANGES IN SLOVENIA FROM THE FRANCISCEAN CADASTER UNTIL TODAY. *Acta Geographica Slovenica-Geografski Zbornik,* 59**,** 63-81.

GEBHART, W. 2011. *"... zur Aufmunterung der Landescultur" : die große franziszeische Katastervermessung und das Herzogtum Salzburg.*

GIMMI, U., LACHAT, T. & BURGI, M. 2011. Reconstructing the collapse of wetland networks in the Swiss lowlands 1850-2000. *Landscape Ecology,* 26**,** 1071-1083.

GOBBI, S., CIOLLI, M., LA PORTA, N., ROCCHINI, D., TATTONI, C. & ZATELLI, P. 2019. New Tools for the Classification and Filtering of Historical Maps. *ISPRS International Journal of Geo-Information,* 8**,** 455.

GODFREY, B. & EVELETH, H. 2015. An Adaptable Approach for Generating Vector Features from Scanned Historical Thematic Maps Using Image Enhancement and Remote Sensing Techniques in a Geographic Information System. *Journal of Map & Geography Libraries,* 11**,** 18-36.

GU, J. X., WANG, Z. H., KUEN, J., MA, L. Y., SHAHROUDY, A., SHUAI, B., LIU, T., WANG, X. X., WANG, G., CAI, J. F. & CHEN, T. 2018. Recent advances in convolutional neural networks. *PATTERN RECOGNITION,* 77**,** 354-377.

HAASE, D., WALZ, U., NEUBERT, M. & ROSENBERG, M. 2005. Changes to Central European landscapes - Analysing historical maps to approach current environmental issues, examples from Saxony, Central Germany. *Land Use Policy,* 24**,** 248-263.

HEROLD, H. 2018. *Geoinformation from the Past : Computational Retrieval and Retrospective Monitoring of Historical Land Use,* Wiesbaden, Springer Spektrum.

HOSSAIN, M. D. & CHEN, D. 2019. Segmentation for Object-Based Image Analysis (OBIA): A review of algorithms and challenges from remote sensing perspective. *ISPRS JOURNAL OF PHOTOGRAMMETRY AND REMOTE SENSING,* 150**,** 115-134.

IGNJATIĆ, J., BAJIC, B., RIKALOVIC, A. & CULIBRK, D. 2018. *Deep Learning for Historical Cadastral Maps Digitization: Overview, Challenges and Potential*.

IOSIFESCU, I., TSORLINI, A. & HURNI, L. 2016. Towards a comprehensive methodology for automatic vectorization of raster historical maps. *e-Perimetron,* 11**,** 57-76.

KERSAPATI, M. & GRAU-BOVÉ, J. 2023. Geographic features recognition for heritage landscape mapping – Case study: The Banda Islands, Maluku, Indonesia. *Digital Applications in Archaeology and Cultural Heritage,* 28**,** e00262.

KUPKOVA, L., BICIK, I. & BOUDNY, Z. 2019. LONG-TERM LAND-USE / LAND-COVER CHANGES IN CZECH BORDER REGIONS. *Acta Geographica Slovenica-Geografski Zbornik,* 59**,** 107-117.

LAND SALZBURG. 2012. *Franciszäischer Kataster ab sofort online verfügbar* [Online]. Salzburg: Land Salzburg. Available: https://service.salzburg.gv.at/lkorrj/Index?cmd=detail_ind&nachrid=49619 [Accessed 09.01.2024].

LECUN, Y., BOSER, B. E., DENKER, J. S., HENDERSON, D., HOWARD, R. E., HUBBARD, W. E. & JACKEL, L. D. Handwritten Digit Recognition with a Back-Propagation Network. Neural Information Processing Systems, 1989.

MA, L., LI, M. C., MA, X. X., CHENG, L., DU, P. J. & LIU, Y. X. 2017. A review of supervised object-based land-cover image classification. *ISPRS JOURNAL OF PHOTOGRAMMETRY AND REMOTE SENSING,* 130**,** 277-293.

MA, L., LIU, Y., ZHANG, X. L., YE, Y. X., YIN, G. F. & JOHNSON, B. A. 2019. Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS JOURNAL OF PHOTOGRAMMETRY AND REMOTE SENSING,* 152**,** 166-177.

MARCEAU, D. J. 1999. The Scale Issue in the Social and Natural Sciences. *Canadian Journal of Remote Sensing,* 25**,** 347-356.

MAXWELL, A. E., WARNER, T. A. & FANG, F. 2018. Implementation of machine-learning classification in remote sensing: an applied review. *INTERNATIONAL JOURNAL OF REMOTE SENSING,* 39**,** 2784-2817.

MENDOZA, F. & LU, R. 2015. Basics of Image Analysis.

O'HARA, R., MARWAHA, R., ZIMMERMANN, J., SAUNDERS, M. & GREEN, S. 2024. Unleashing the power of old maps: Extracting symbology from nineteenth century maps using convolutional neural networks to quantify modern land use on historic wetlands. *Ecological Indicators,* 158**,** 111363.

PETIT, C. C. & LAMBIN, E. F. 2002. Impact of data integration technique on historical land-use/land-cover change: Comparing historical maps with remote sensing data in the Belgian Ardennes. *Landscape Ecology,* 17**,** 117-132.

RIBEIRO, D. & HRIBAR, M. S. 2019. ASSESSMENT OF LAND-USE CHANGES AND THEIR IMPACTS ON ECOSYSTEM SERVICES IN TWO SLOVENIAN RURAL LANDSCAPES. *Acta Geographica Slovenica-Geografski Zbornik,* 59**,** 143-159.

SERGEY, I. & CHRISTIAN, S. 2015. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. PMLR.

SHIVAKUMAR, B. R. & NAGARAJA, B. G. 2023. Expert System Classifier for RS Data Classification.

STATUTO, D., CILLIS, G. & PICUNO, P. 2017. Using Historical Maps within a GIS to Analyze Two Centuries of Rural Landscape Changes in Southern Italy. *Land,* 6**,** 15.

TRIMBLE. 2024a. *Classification using the Template Matching Algorithm* [Online]. Available: https://docs.ecognition.com/v9.5.0/eCognition_documentation/User%20Guide%20Developer/6%20About%20Classification.htm#TemplateMatching [Accessed 26.02.2024].

TRIMBLE. 2024b. *Deep Learning (CNN) Algorithms* [Online]. Available: https://docs.ecognition.com/eCognition_documentation/Reference%20Book/02%20Algorithms%20and%20Processes/9%20Deep%20Learning%20(CNN)%20Algorithms/Deep%20Learning%20(CNN)%20Algorithms.htm [Accessed 21.02.2024].

TRIMBLE. 2024c. *Trimble eCognition* [Online]. Available: https://geospatial.trimble.com/en/products/software/trimble-ecognition [Accessed 10.01.2024].

TRIMBLE n.d. Tutorial 7 — Convolutional Neural Networks in eCognition

UHL, J., LEYK, S., CHIANG, Y.-Y., DUAN, W. & KNOBLOCK, C. A. 2018. Spatialising uncertainty in image segmentation using weakly supervised convolutional neural networks: a case study from historical map processing. *IET Image Processing,* 12**,** 2084-2091.

UHL, J., LEYK, S., CHIANG, Y.-Y. & KNOBLOCK, C. 2022. *Towards the automated large-scale reconstruction of past road networks from historical maps*.

UHL, J., LEYK, S., DUAN, W., LI, Z., SHBITA, B., CHIANG, Y.-Y. & KNOBLOCK, C. 2021a. Towards the large-scale extraction of historical land cover information from historical maps. *Abstracts of the ICA,* 3**,** 1-2.

UHL, J., LEYK, S., LI, Z., DUAN, W., SHBITA, B., CHIANG, Y.-Y. & KNOBLOCK, C. 2021b. *Combining Remote Sensing Derived Data and Historical Maps for Long-Term Back-Casting of Urban Extents*.

ULLOA, Y., STAHLMANN, R., WEGMANN, M. & KOELLNER, T. 2020. Over 150 Years of Change: Object-Oriented Analysis of Historical Land Cover in the Main River Catchment, Bavaria/Germany. *Remote Sensing,* 12**,** 4048.

WIKIMEDIA COMMONS 2014. Legende zu den Kartenblättern des Franziszeischen Katasters. Wikimedia Commons.

WOKRIE 2015. Ausschnitte aus dem Sattlerpanorama. *In:* SATTLERPANORAMA-0006 (ed.). Wikimedia Commons.

XIA, X., HEITZLER, M. & HURNI, L. 2022. CNN-BASED TEMPLATE MATCHING FOR DETECTING FEATURES FROM HISTORICAL MAPS. *XXIV ISPRS CONGRESS IMAGING TODAY, FORESEEING TOMORROW, COMMISSION II.*

ZATELLI, P., GABELLIERI, N. & BESANA, A. 2022a. Digitalization and Classification of Cesare Battisti's Atlas of 1915. *ISPRS International Journal of Geo-Information,* 11**,** 238.

ZATELLI, P., GOBBI, S., TATTONI, C., LA PORTA, N. & CIOLLI, M. 2019. OBJECT-BASED IMAGE ANALYSIS FOR HISTORIC MAPS CLASSIFICATION. *ISPRS - International Archives of the Photogrammetry Remote Sensing and Spatial Information Sciences,* XLII-4/W14**,** 247-254.

ZATELLI, P., TATTONI, C., GOBBI, S., CANTIANI, M. G., LA PORTA, N. & CIOLLI, M. 2022b. MODELING OF FOREST LANDSCAPE EVOLUTION AT REGIONAL LEVEL: A FOSS4G APPROACH. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.,* XLVIII-4/W1-2022**,** 553-560.

Appendix

Part I: Classification result of the main study area around Salzburg.

Part II: Classification result of the High Tauern study area.

Land Use and Land Cover around the City of Salzburg, 1830

This map shows land cover and land use types around the city of Salzburg extracted from the Franciscean Cadastre. Mapping for the Province of Salzburg took place between 1823 and 1830 in a scale of 1:1440 for urban areas and 1:5760 for rural regions.
For this classification, classes were summarized and simplified based on their spectral properties.

LULC Classes

- Agriculture
- Forest
- Garden
- Meadow
- Other
- Path
- Footpath
- Chaussee
- Buildings
- Commercial Buildings
- Water

Creator: Simon Meyer
Contact: S1039004@stud.plus.ac.at

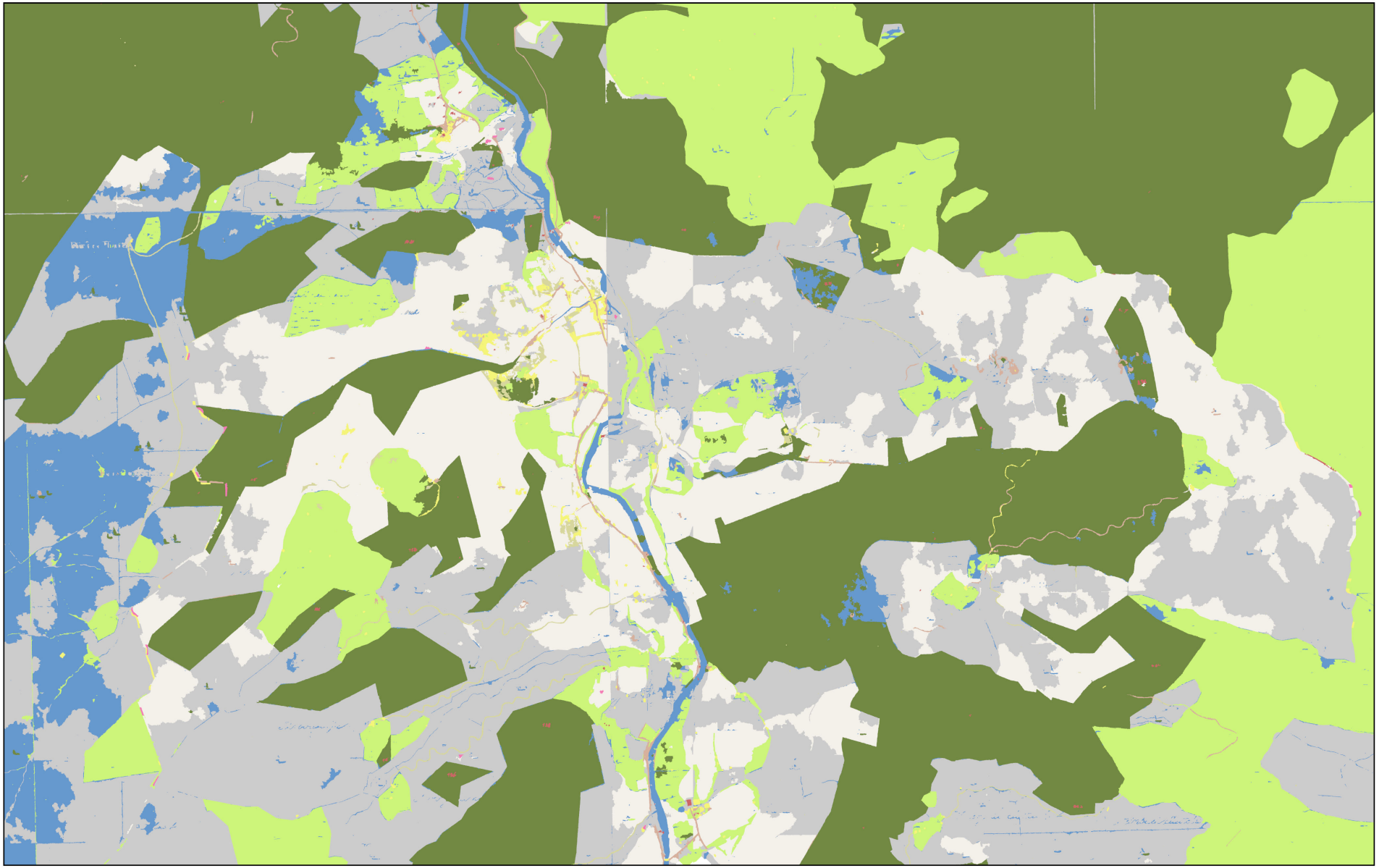
Credits: SAGIS

Spatial Reference
Name: MGI Austria GK M31
PCS: MGI Austria GK M31
Datum: MGI
Projection: Transverse Mercator
WKID: 31258

Scale: 1:15.000

0 250 500 Meters

Land Use and Land Cover between Rauris and Taxenbach, 1830



Scale: 1:15.000

0 250 500 Meters

Spatial Reference
Name: MGI Austria GK M31
PCS: MGI Austria GK M31
Datum: MGI
Projection: Transverse Mercator
WKID: 31258

Creator: Simon Meyer
Contact: S1039004@stud.plus.ac.at

Credits: SAGIS

This map shows land cover and land use types around the city of Salzburg extracted from the Franciscean Cadastre. Mapping for the Province of Salzburg took place between 1823 and 1830 in a scale of 1:1440 for urban areas and 1:5760 for rural regions.
For this classification, classes were summarized and simplified based on their spectral properties.