IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received April 6, 2021, accepted April 19, 2021, date of publication April 22, 2021, date of current version April 30, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3074897

Automatic Detection of Road Types From the Third Military Mapping Survey of Austria-Hungary Historical Map Series With Deep Convolutional Neural Networks

YEKTA SAID CAN, PETRUS JOHANNES GERRITS, AND M. ERDEM KABADAYI

College of Social Sciences and Humanities, Koç University, 34450 Istanbul, Turkey

Corresponding author: Yekta Said Can (ycan@ku.edu.tr)

This work was supported by the European Research Council (ERC) Project: Industrialisation and Urban Growth from the mid-nineteenth century Ottoman Empire to Contemporary Turkey in a Comparative Perspective, 1850–2000 under the European Union's Horizon 2020 Research and Innovation Program Grant (acronym UrbanOccupationsOETR) under Agreement 679097.

ABSTRACT With the increased amount of digitized historical documents, information extraction from them gains pace. Historical maps contain valuable information about historical, geographical and economic aspects of an era. Retrieving information from historical maps is more challenging than processing modern maps due to lower image quality, degradation of documents and the massive amount of non-annotated digital map archives. Convolutional Neural Networks (CNN) solved many image processing challenges with great success, but they require a vast amount of annotated data. For historical maps, this means an unprecedented scale of manual data entry and annotation. In this study, we first manually annotated the Third Military Mapping Survey of Austria-Hungary historical map series conducted between 1884 and 1918 and made them publicly accessible. We recognized different road types and their pixel-wise positions automatically by using a CNN architecture and achieved promising results.

INDEX TERMS Convolutional neural networks, digital humanities, digital preservation, document analysis, geospatial analysis, geospatial artificial intelligence, road type detection, image processing.

I. INTRODUCTION

Historical documents are precious cultural sources that help researchers investigate social, historical and economic perspectives of the past. The digitization of them grants direct access to researchers and the public. Nevertheless, due to maintenance purposes, direct access to these archives could be restricted or not possible. With the help of increased digitization processes in the last decades, they can be analyzed, and researchers can retrieve new information. There are also organized and well-funded efforts to make digitized and georeferenced historical maps publicly available including detailed metadata [1]. Recently, automatic processing techniques are applied to the historical maps for information extraction. Road types, intersection of roads, human settlements, forest and flora cover information were extracted in these studies. For many academic disciplines, these maps are essential in making long-term spatial quantitative analyses and assessments, primarily due to the wide variety of annotated historical and geographical information they contain. They offer one of the most comprehensive ways, if not the only way, in which historians, archaeologists and geographers can gain insight into how historical transport networks developed over time. Often, however, we can only find these historical maps as scanned archival documents, making the process of extracting annotated information very time-consuming and complex. The digitization and manual annotation of 300.000 km of road features with more than 64,000 segments took 1250 hours with two graduate students. Training and testing the model lasted for 7 days maximum on GPU. As far as the complexity is concerned, our model has 32.8M parameters. When we searched the literature for historical map processing, the training times and complexity of models are not generally reported but Uhl *et al.* [2] reported that 138 million parameters used

The associate editor coordinating the review of this manuscript and approving it for publication was Xiaochun Cheng.

in their VGGNet-16 models. Our model has relatively lower number of parameters, training time and complexity because of the pretrained architectures.

CNN architectures outperformed feature extraction-based models in the image processing field and solved many problems [8]. However, they require a large amount of annotated data for training. For historical maps, this requires an unprecedented scale of manual data entry and data annotation which makes it challenging to use CNN models on historical maps. Towards this effort, in this research, we would like to contribute by making a large-scale public database with annotated and/or labeled historical map features available. This dataset can reproduce the shared research, as well as enable research towards CNN training on historical maps. The links to the repository are shared later in this paper.

The map series used in this research, the Third Military Mapping Survey of Austria-Hungary or *Generalkarte von Mitteleuropa* in German (hereafter Generalkarte),' was created between 1884 and 1918 by the Austro–Hungarian Military Geographical Institute in Vienna. The map is also known as the Third Mapping Survey of Austria-Hungary. This survey was the result of the third, the last and cartographically the most accurate Austro–Hungarian imperial surveying effort with the massive additional geographical scope of mapping Central and East Europe in an unprecedented topographical detail and precision between 1884 and 1918 [9]. The map series is documented in 267 individual map sheets that cover the complete Austro–Hungarian Empire, with the outer corners laying in as far as Cologne, Istanbul, Nizza and Kiev. The historical map has been through many iterations and was republished in 50 separate series since its initial release between 1886 and 2000 [10]. Furthermore, the transport facilities were of utmost importance to the surveyors of the Generalkarte due to military purposes [11]. Therefore, due to the size, historical impact and spatio-temporal accuracy, we consider focusing on this particular map series essential for automatic feature recognition.

This research provides a novel approach for extracting detailed semantic information of historical maps using Deep Convolutional Neural Networks. In time, we are convinced that CNNs will be able to accomplish the complicated task of automating this digitization process. This research is aimed towards this goal by making use of a large-scale manually digitized transport network that was supervised and extracted by hand from the Generalkarte. For preparing ground truths of our dataset for training a CNN, we digitized a historical transport network. The digitized historical network was initially used for the development of a spatial interaction model to reconstruct a multimodal transport infrastructure for Southeast Europe for around 1900. We previously constructed a preliminary multimodal transport network for the entire territory of the Ottoman Empire in 1899 following the same method of manual annotation. However, that model had only one type of railroad and two types of roads [12]. Generalkarte, on the other hand, has five types of railroads and 15 types of roads, as can be seen in the legend of the map in Figure 3. We re-used the transport model and converted the transport road segments that were digitized and used them as ground truth datasets in our Deep CNN because the annotated features it contains can also serve as an excellent training and verification data set for automatized feature extraction. The contributions of our study are three-fold:

- We created the first annotated historical map dataset for road type detection. It includes more than 7000 images and their labels and we made it publicly accessible.
- We applied CNNs to the Generalkarte historical map series for road type detection with the most successful pretrained network [4] (Resnet50) in historical map processing.
- We also explored the performance of Unet architecture as the pretrained network for road type detection which has not been used in historical map processing studies.

The rest of the paper is organized as follows. The literature on historical map processing is provided in Section 2. The description of the map series and the dataset is presented in Section 3. We described the methodology in Section 4. In Section 5, the experimental results and discussion are presented. We summarize our findings and mention our planned future works in the concluding Section 6.

II. RELATED WORKS

Recently, more and more studies have tried to extract information automatically from historical maps with the advance of computer vision algorithms. The researchers tried to make sense out of these maps and make use of the retrieved information. There are many significant challenges for automatically extracting features from raster maps in geographical information systems (GISs). The first challenge is that elements in the historical maps might overlap with each other such as road lines, elevation contour lines, marks or soil features and populated place names [6]. Secondly, the digitized map could have low quality due to scanning or compression processes, which could further complicate the image recognition task. Lastly, some old maps may be degraded that makes the development of an automatic procedure more challenging.

Several authors proposed automatic techniques to extract geographical features in scanned thematic historical maps. Some example maps used for feature recognition can be listed as the topographic maps of the United States Geological Survey (USGS), a historical map of France from the 19th century [6] and the Swiss topographic maps (Siegfried maps) [7] (see Table 1). USGS provided maps are the most commonly used historical maps for feature extraction.

Uhl *et al.* used USGS maps for human settlement footprint retrieval in their studies [2], [3]. By applying a weakly supervised CNN, they achieved promising results (i.e., recall of up to 0.96, F_measure of up to 0.79). Saeedimoghaddam and Stepinski [5] developed a road intersection point detection system by using USGS historical maps. They achieved a 0.8 F_measure with a fast CNN algorithm. Railroad features were also studied in the literature. Chiang *et al.* [4] presented

TABLE 1. The comparison of our study with the historical map feature extraction studies on different datasets.

Study	Dataset	Dataset Type	Method	Performance
[3][2017]	USGS	Human Settlement Footprint	CNN + LeNet50	0.11 F_measure - Individual Buildings
[4][2020]	USGS	Railroad Features	CNN + ResNet	0.622 IoU
[5] [2020]	USGS	Road Intersection Points	CNN (Fast)	0.8 F_measure
[2] [2020]	USGS	Human Settlement Pattern	Weakly Supervised CNN	0.18 - 0.42 F_measure 0.28 - 0.65 IoU - 19th-century maps
[6] [2013]	Historical maps of France	Forest Features	K means + Color Space Conversion	0.9 Kappa Score
[7] [2016]	Swiss topographic map	Forest Cover Information	N/A	N/A
HistMapCNNRes50 [2021]	Generalkarte	Road Type Information	CNN + Resnet50	0.4123 F_measure, 0.4649 IoU
HistMapCNNUnet [2021]	Generalkarte	Road Type Information	CNN + Unet	0.5321 F_measure, 0.4644 IoU

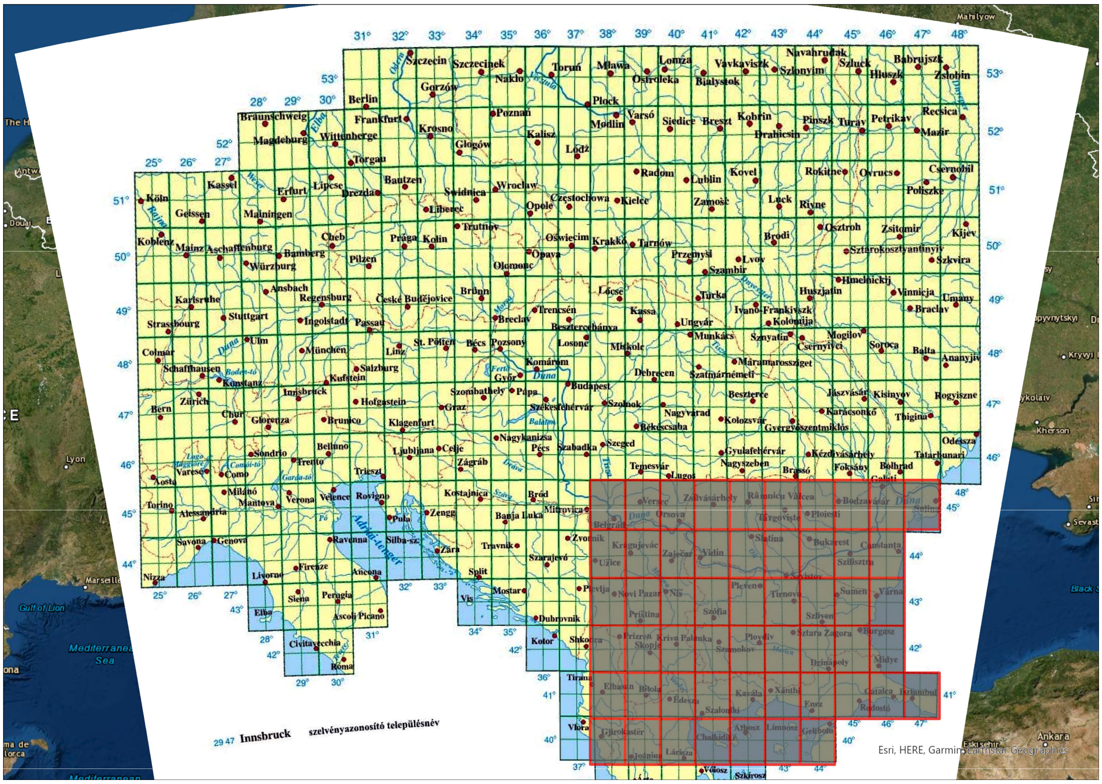

FIGURE 1. Coverage of study area (in red) and the full generalkarte extent [13].

a thorough experiment for extracting railroad features from USGS historical topographic maps as a case study. They achieved a 23.09 IoU score with CNN + Resnet50 architectures. Forest features were extracted from a historical map of France and a Swiss topographic map (Siegfried maps). Herrault *et al.* [6] achieved a 0.9 Kappa score for extracting forest features by using an unsupervised K-means algorithm and color space conversion. Duan *et al.* [15] presented a method that automatically aligns label data with geographic properties of historical maps. They created ground truth labels automatically, but the alignment of these labels with geographic map features is also a challenge. They proposed a generalizable framework for the alignment of these ground truth vectors. They achieved 100% correctness and 20% completeness for river and railroad detection problems. They improved the alignment technique in a recent study [16]. They used reinforcement learning (RL) for aligning automatically extracted ground truth data and map features. RL improves especially the completeness results for detecting railroads and rivers. Template matching from historical maps was also tried in the literature. Budig and van Dijk [17] applied active learning for detecting and finding positions of meta-data images such as text or place markers from historical maps. They achieved 80% for detecting nine image templates. They further investigated smart crowdsourcing for extracting building footprints from historical maps in another study [18] to increase the annotated data. They developed an algorithm for combining crowdsourcing answers for building footprint detection and achieved better accuracy when compared to individual answers.

As can be seen from the literature, the most important challenge is to obtain annotated data. To overcome scarce annotated data problems, crowdsourcing, automatic annotation and alignment techniques, weakly supervised CNN algorithms were employed. This shows the importance of annotated historical maps for improving state of art. Our study is the first one to annotate and retrieve information from the Generalkarte historical map series which would reveal very substantial information on the transport infrastructure in Southeast Europe before and during World War I. We also applied state-of-the-art CNN architectures to this map series for detecting seven types of roads. Although there are several studies that extract features from different historical maps by using machine learning techniques, our study is the most comprehensive road analysis by using machine learning in the literature.

III. DATASET DESCRIPTION

The digitized geospatial transport dataset used in this research as ground-truth labels were manually created by the UrbanOccupationsOETR project based on the Generalkarte. The collection of transport features can be regarded as the oldest topographical data set from this region that is derived from historical maps. It was initially used to gain a better understanding of historical transport routes and estimating reconstructing detailed road elevation profiles. The digital data set was created by geographically locating and referencing the individual map sheets, commonly referred to as georeferencing. After georeferencing the historical map, vector-based polylines were drawn manually over the corresponding topographical and geographical elements. On the whole, the Generalkarte map collection covers a majority of Central and East Europe in a total of 267 map tiles. However, for this research, we made use of 54 map tiles that cover modern Bulgaria and large parts of Greece, the Republic of North Macedonia, Serbia and Albania (see Figure 1). The necessity for deep learning for automatic feature retrieval of road features can be showcased very well by the time and effort it took to digitize this historical map manually. Between 2017 and 2019, it took two research fellows more than 1250 hours to digitize 300.000 km of road features with more than 64,000 segments based on this historical map. Creating a deep learning model can significantly reduce the time spent on creating such networks and it is our hope that this research can be transferred and used to analyze other areas and different periods.

Our choice to focus on the Generalkarte is not only based on it being one of the oldest historical map collections with accurate transport features, but also because of the rich annotations represented on the map (see Figure 2 and 3). The Austro–Hungarian cartographers were very explicit in documenting the large variety of topographical and geographical features. This rich annotation can provide valuable information as it gives a direct and detailed link with the historical conditions of the study area at that time. Table 2 shows the presence of the individual road types on the historical maps,

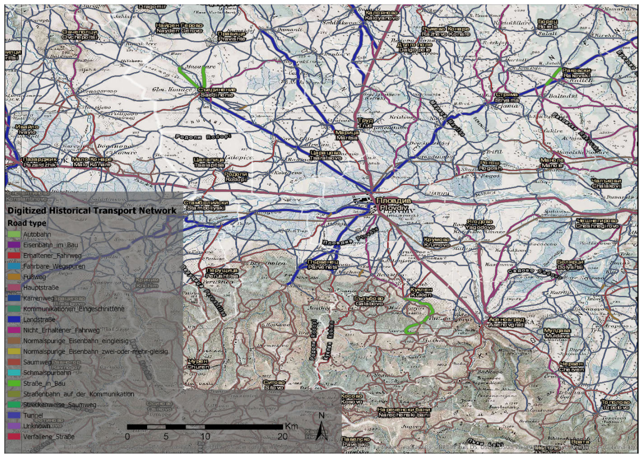

FIGURE 2. Generalkarte labeled example upon modern roads from Plovdiv, Bulgaria [13].

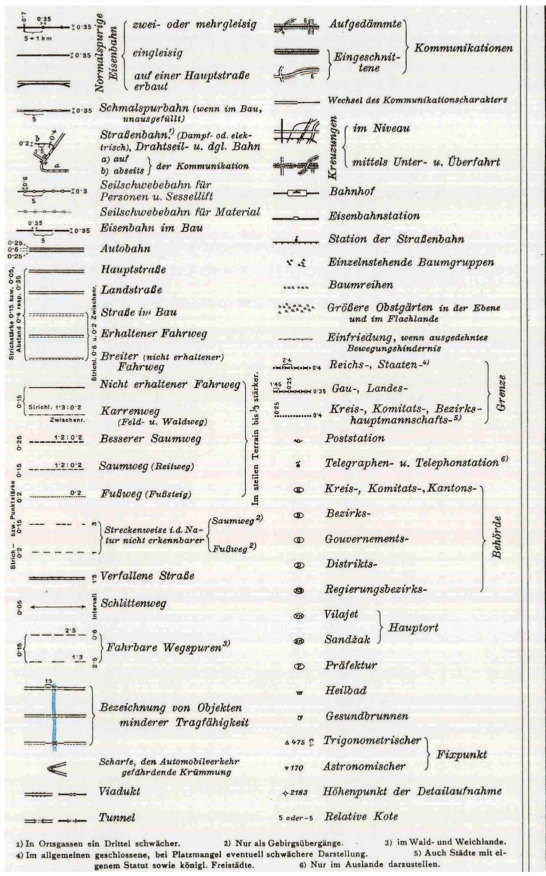

FIGURE 3. Legend of the generalkarte and variety of roads [14].

which were computed after digitizing. We divided the training and test sets by using the map tiles. The training and test sets have an equal number of map tiles, but they sometimes have a different number of images for training and test of some road types. For this research, we used the following road types for analysis (see Figure 3 for the legend of the map):

- 1) bridle paths (German: *Saumweg*),
- 2) cart roads (German: *Karrenweg*),

TABLE 2. Different road types are included in the analysis and their appearance on the historical map as segment count. N. stands for Nicht. Train. stands for training and Img. stands for images.

Road Type	# of Train. Img.	# of Test Img.	Total
Karrenweg	9639	8957	18596
Saumweg	9255	8821	18076
N. erhaltener Fahrweg	4844	4682	9526
Landstrasse	4074	3309	7383
Erhaltener Fahrweg	3039	3278	6317
Hauptstrasse	1437	954	2391
Fussweg	828	919	2391

- 3) main roads (German: *Hauptstrasse*),
- 4) footpaths (German: *Fussweg*,
- 5) country roads (German: *Landstrasse*)
- 6) maintained and 7) non-maintained roads (German: *Erhaltener Fahrweg* and *Nicht erhalterner Fahrweg* respectively).

As a public data set, we created label and image datasets containing all features, which were automatically organized into sub-folders using a python script. The output images were converted from a tif file format to png to ensure compatibility with the deep learning training module. The dataset contains both a dataset for training as well as testing purposes so that our research can be recreated. Labels were stored as CNN_Model_RCNN_PNG\labels\roadType\label.png. The dataset was controlled manually and cleaned. Each road type has 1000 images (256 × 256) and their corresponding labels. Half of these images and labels are placed into the training folder, and the remaining 500 of them are put into the test folder. The dataset can be accessed at https://urbanoccupations.ku.edu.tr/publicdatasets/. We created one downloadable dataset that contains both the training and test sets of images and labels.

IV. METHODOLOGY

Methodologically we separated this research into two steps. The first step involved the creation and preparation of the labeled features in a GIS software environment. This is needed for our labeled dataset and validating our deep learning model. For creating suitable labeled information, we utilized several geoprocessing tools, modules and python environment available in the ArcGIS Pro software package [19], as well as GDAL [20], an open-source python library for geospatial data analysis. The geospatial data is divided into training and test data in order to validate and understand the accuracy of our input model (see Figure 4). The second computational component of this research makes use of

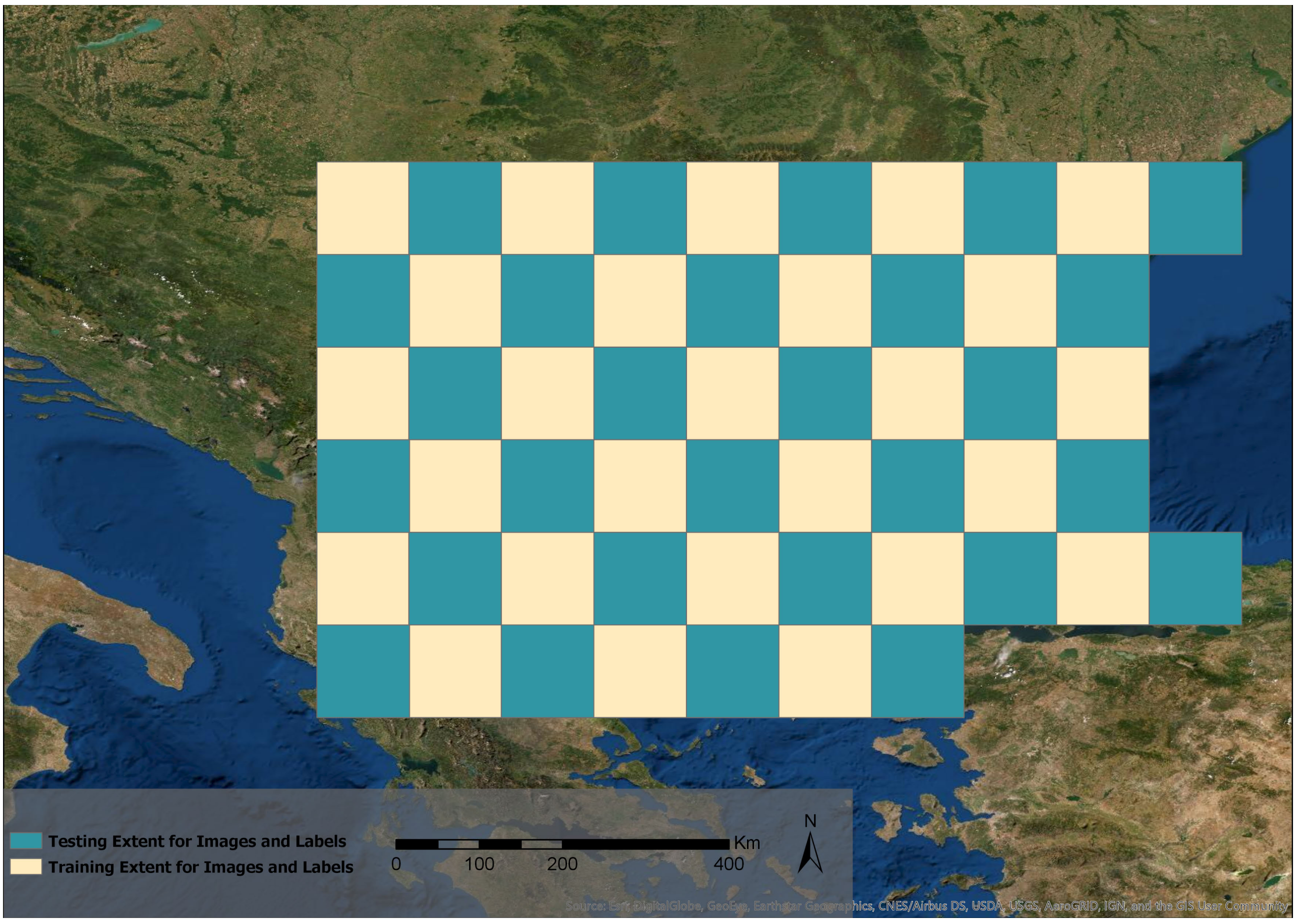

FIGURE 4. The division of training and test data displayed geographically over the study area.

Tensorflow [21] for the creation of our Deep Learning Model, a free and open-source software library for data flow and differentiable programming across a range of tasks, often used for machine learning applications such as neural networks.

A. GEOSPATIAL GROUND-TRUTH DATA SET

1) PROCESSING OF HISTORICAL IMAGERY

The individual map sheets were georeferenced using a Projected Coordinate System (Europe Lambert Conformal Conic, ESRI: 102014) suitable for the entire research area. The 54 individual map sheets were separately georeferenced before digitizing. The scanned and georeferenced maps had an image resolution of 4644 by 4097 and dpi of 96. The maps were not preprocessed prior to creating the labeled dataset and were used in their original form.

2) CREATION OF THE GROUND-TRUTH LABELS OF THE GEOSPATIAL DATASET

The geospatial label dataset was created in ArcGIS Pro software by manually drawing polyline features over georeferenced historical maps. Each road type was annotated separately, allowing a more detailed reconstruction. In order to reshape geographic polylines into labeled features usable for the deep learning CNN model, the first step was to convert polylines to polygons. We determined the width of the new polygons by making sure that they completely covered underlying road features from the raster-based historical map. We showed a representation of the outcome in Figure 2. Transforming the line segments into polygons was done by using a buffer analysis ArcGIS Pro and using different width fields per sub-classes.

After creating the polygon-based labeled dataset, the next step involved the preparation for the extraction of images and labels for the Deep Learning model. Using the "Export training data for Deep Learning" geoprocessing tool in ArcGIS Pro, we converted our polygon features to RCNN masked images and labels. The road types were automatically assigned to a sub-folder structure and necessary metadata was provided as well. Labels and images were divided into image masks with a width of 256 by 256 pixels, with a stride overlap of 128 by 128 pixels. The masked images were exported as a png image file type; the labels could only be exported as tif files. In order for our Deep Learning Training system to work, we converted the images from tif to png files as well, using GDAL and a python script for automating the workflow.

B. TRAINING DEEP LEARNING CNN MODEL

We used the open-source dhSegment toolbox [23] as in our previous segmentation studies [24] and [25]. The authors describe the toolbox as a general and flexible architecture for pixel-wise segmentation related tasks on historical documents. This toolbox is comprised of two components. The first component is a Fully Convolutional Neural Network (FCNN). The scanned and digitized image of a document is provided to the FCNN (the toolbox does not limit the size of the image if there is enough memory to process it). The toolbox outputs the probability map of pixels generated by the ReLU function that belongs to the trained object types. This map could be used to transform the prediction map to the desired output in the post-processing step. We provided an overview of the system in Figure 5.

The network architecture has two paths: an expanding and contracting path, namely. These paths formed a Unet architecture [26]. The contracting path has repeated convolutional layers that reduce spatial information and increase feature information. Conversely, the expanding path merges the feature and spatial information with a sequence of up-convolutions and concatenations through high-resolution features obtained from the contracting path. The contracting path follows the structure of the deep residual network ResNet-50 [27] (it is not the same because of memory efficiency reasons). The ResNet-50 architecture is demonstrated in yellow blocks. The number of feature channels is limited

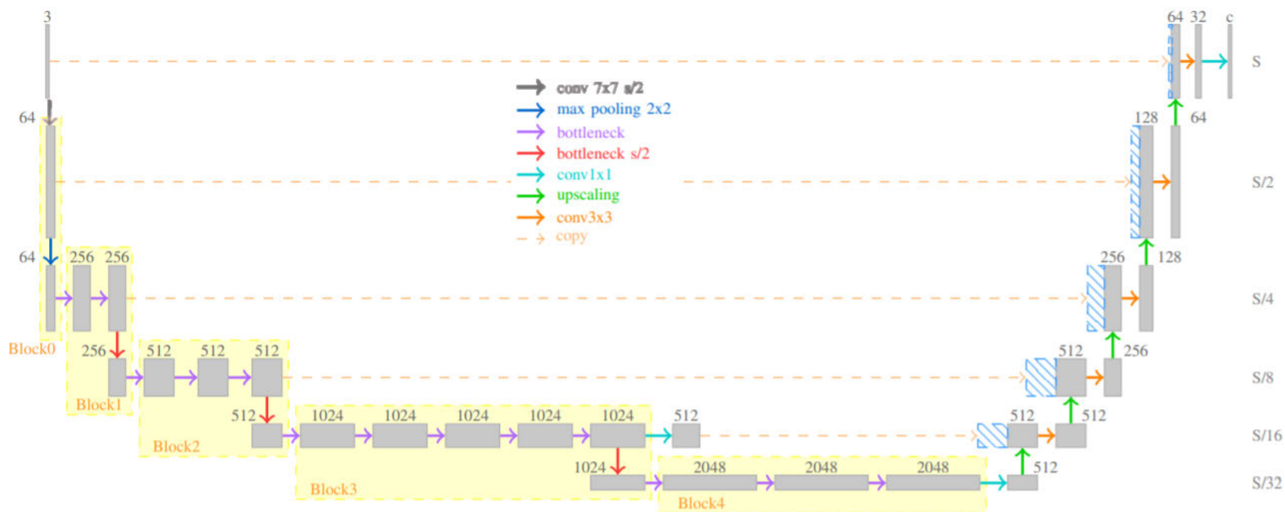

FIGURE 5. The architecture of the dhSegment. It forms a unet architecture. It was adapted from [22], p. 9, ©2019 IEEE.

to 512 in the expansive path to restrict the number of training parameters, thus decreasing the contracting path (light blue arrows). Expanding steps double the size of the feature map and halves the feature channel number. The output prediction has an equal size with the input image, and the number of feature channels forms the required number of classes. This last convolution layer is parameterized as follows: c (number of classes) filters, 3×3 kernel, stride and padding of 1. It is followed by a softmax layer that calculates the class probabilities of each pixel.

The expanding path is comprised of five blocks plus a final convolutional layer which assigns a class to each pixel. Each deconvolutional step is comprised of an upscaling of the former block feature map, a concatenation of the upscaled feature map with a copy of the corresponding contracting feature map and a 3×3 convolutional layer followed by a rectified linear unit (ReLU). The number of feature channels in step $i = 4$ and $i = 5$ is diminished to 512 by a 1×1 convolution before concatenation for reducing the number of parameters and memory usage. The upsampling is achieved by applying a bilinear interpolation.

The input images are map fragments that have 256×256 dimensions. Paths use pretrained weights from a general image classification task where the system learns high-level features (ImageNet). These weights increase robustness and improve generalization. The system's architecture has 32.8M of parameters, but only 9.36M parameters are required to be trained because of the pretrained weights in the contracting path. With the pretrained weights in the network, the training time was decreased substantially [23]. We used GPUs for training the models.

For training the models, L2 regularization is used with 10-6 weight decay [23]. We use a learning rate with an exponential decay rate of 0.95 and an initial value in [10-5; 10-4]. Xavier's initialization [28] and Adam optimizer [29] were employed. Batch renormalization [30] was used to evade the lack of diversity issue. The toolbox also downsized images and divided them into 300 to 300 patches to fit the memory better and to provide batch training. By adding the margins, the toolbox prevents border effects. The training process uses different data augmentation methods such as rotation (from -0.2 to 0.2 rad), scaling (coefficient from 0.8 to 1.2) and mirroring. The output of the neural network can be used in post-processing steps. We trained dhSegment on our data with 100 epochs since the model is pre-trained and converges fastly. We used mini-batches of size 5.

V. EXPERIMENTAL RESULTS AND DISCUSSION

In this section, we provided the results for detecting the selected seven different road types. We used two different pretrained architectures, namely Unet and Resnet50. We first describe the metrics and provide the results, and discuss them.

A. PREPARING THE DATASET FOR EVALUATION

We trained 14 different models for evaluating the performance of our road type detection (seven different road types) system. Half of these models were trained by using pretrained Resnet50 architecture, and the other half were trained by using a pretrained Unet architecture. These architectures were selected among the commonly applied networks such as Vgg16, GoogleNet, Resnet50 and Unet. Resnet50 is reported to have better IoU results than Vgg16 and GoogleNet when applied to the historical maps [4]. We selected Unet architecture because it is relatively new and it is not applied in historical map processing studies before. We balanced the number of images in the training and test sets by removing extra samples from the majority sets. Map images were divided into 50% training and 50% test parts in this way. Binary classification models were formed for each road type (road versus background).

B. METRICS

To assess the performance of our road detection scheme, five metrics were employed. These metrics are pixel-wise classification accuracy, Intersection over Union, pixel-wise precision, recall and F_measure. They are widely used in different image analysis applications for detecting objects [31]. We will briefly describe these metrics in this section.

1) PIXEL-WISE CLASSIFICATION ACCURACY

The first metric is pixel-wise accuracy. It can be calculated by dividing the accurately classified pixels in all documents by the number of all pixels (for all object types).

2) INTERSECTION OVER UNION

We also calculated the Intersection over Union (IoU) metric. The segmented objects' actual area can be called the ground truth, whereas the connected areas are formed by connecting the adjacent pixel classifications that belong to the same class can be called the prediction area. The IoU can be computed by dividing the intersection of these two areas into the union of these areas.

3) PIXEL-WISE PRECISION, RECALL AND F_measure

We calculated and used the pixel-wise precision, recall and F_measure metrics. They are computed for all pixels on the documents (for all road types). Precision and recall metrics are used to calculate the F_measure. It is widely used in the cases where there is a class imbalance to present classification performance in more detail. Their formulas are provided as:

$$Precision = \frac{TruePositive}{TruePositive + FalsePositive} \tag{1}$$

$$Recall = \frac{TruePositive}{TruePositive + FalseNegative} \tag{2}$$

$$F_{measure} = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{3}$$

C. ROAD TYPE DETECTION RESULTS

For each road type, we trained a CNN model and tested the performance of these models on the test sets. We used

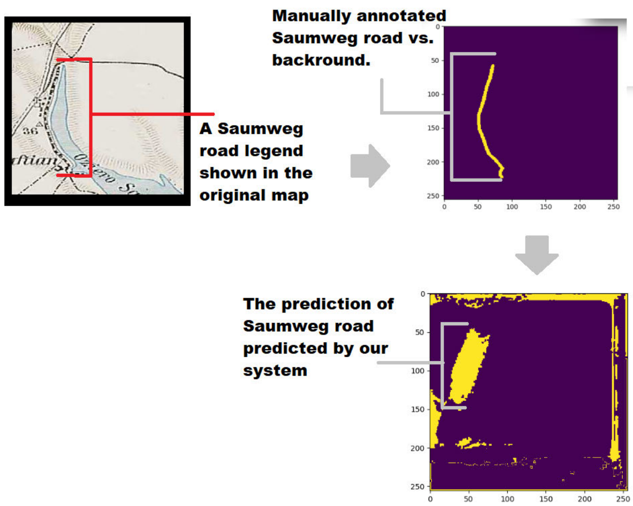

FIGURE 6. A sample prediction of our system. In the top left, the original map; in the top right, the manual label that shows the road on the map. We can see the binary prediction of our system at the bottom.

TABLE 3. IoU and pixel-wise accuracy results obtained by using the pretrained Resnet50 architecture are provided.

Road Type	IoU	Pixel-wise Accuracy
Saumweg	0.4649	0.9299
Karrenweg	0.4388	0.8602
Erhaltener Fahrweg	0.4351	0.8521
Nicht erhaltener Fahrweg	0.4293	0.8069
Landstrasse	0.4288	0.8134
Hauptstrasse	0.4363	0.8320
Fussweg	0.4446	0.8562

TABLE 4. IoU and pixel-wise accuracy results obtained by using the pretrained unet architecture are provided.

Road Type	IoU	Pixel-wise Accuracy
Saumweg	0.4644	0.9288
Karrenweg	0.4378	0.8758
Erhaltener Fahrweg	0.4339	0.8677
Nicht erhaltener Fahrweg	0.4346	0.8693
Landstrasse	0.4299	0.8599
Hauptstrasse	0.4377	0.8753
Fussweg	0.447	0.8941

pretrained Resnet-50 and Unet models for each road type and extracted results. The obtained results are presented with four metrics in Tables 3,4,6, and 7. A sample prediction output

TABLE 5. Precision, recall and F_measure results obtained by using the pretrained Resnet50 architecture are provided.

Road Type	Precision	Recall	F_measure
Saumweg	0.2481	0.1198	0.1616
Karrenweg	0.2872	0.5689	0.3818
Erhaltener Fahrweg	0.281	0.7736	0.4123
Nicht erhaltener Fahrweg	0.0518	0.3359	0.898
Landstrasse	0.0262	0.0569	0.0359
Hauptstrasse	0.1076	0.0315	0.0488
Fussweg	0.0279	0.8165	0.0539

TABLE 6. Precision, recall and F_measure results obtained by using the pretrained unet architecture are provided.

Road Type	Precision	Recall	F_measure
Saumweg	0.2206	0.2262	0.2234
Karrenweg	0.3626	0.9996	0.5321
Erhaltener Fahrweg	0.3639	0.9995	0.5336
Nicht erhaltener Fahrweg	0.0646	1	0.1214
Landstrasse	0.065	0.9946	0.1221
Hauptstrasse	0.079	0.1449	0.1022
Fussweg	0.2813	0.0863	0.132

of our system, the input image and the ground truth labels are shown in Figure 6. Visualization results for a sample road type are demonstrated in Figure 7. Using a pretrained Unet architecture achieves better results when compared to the Resnet50 architecture which shows that it is more suitable for these types of historical map data. Furthermore, the best accuracies are obtained when detecting *Saumweg* and *Karrenweg* and *Erhaltener Fahrweg* road types in terms of F_measure, IoU and pixel-wise accuracy. These roads are frequently seen, and their number of images is higher than the other road types. Their detection accuracies are then higher. Furthermore, CNN architectures require a vast amount of data for training. Although we have one of the most extensive datasets for historical maps, they need more training data for some road types. Another important factor that affects the classification performance is the representation of road legends. Some road type legends could not be distinguished from the background easily due to the thickness, color, style. Therefore, for these road types, the F_measure drops to 0.10, which is a relatively low score. However, it is comparable to the results seen in the literature for processing historical maps.

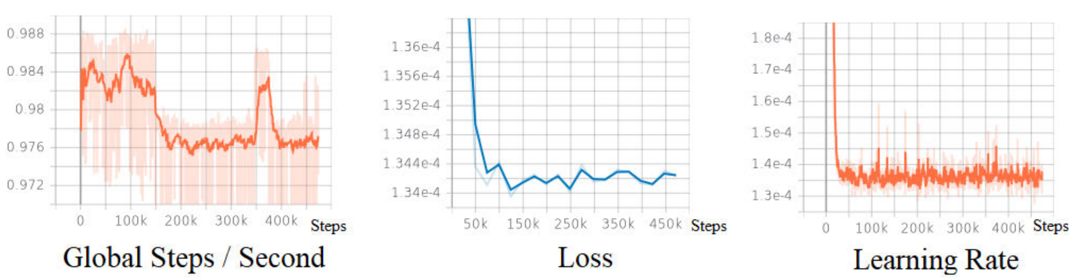

FIGURE 7. Visual results for detecting Saumweg road type with unet pretrained architecture. The left figure shows the global steps per second, the middle figure shows loss and the right figure shows the learning rate with changing steps as a horizontal axis.

When we compare our work with the segmentation techniques applied to different historical maps, our results were aligned with the best-reported results in the literature (see Table 1 for comparison with railroad feature extraction study [3]). Note that the results should be compared with other studies that worked on historical maps in a similar era. When the maps become more recent, the accuracies increase due to the increase in the quality of the maps. Furthermore, because the different techniques were tried in different datasets created for each particular study in the literature, one could not infer the success of a technique over others.

VI. CONCLUSION

In this study, first, the Generalkarte historical map series are manually annotated. The different road types and their pixel-wise positions are detected automatically by using a CNN classifier. We achieved promising results for frequently seen road types such as *Saumweg*, *Karrenweg* and *Erhaltener Fahrweg*. For these road types, we achieved comparable results (around 0.5 F_measure, 0.45 IoU and 0.93 pixel-wise accuracy) with the literature for automatic processing maps of the 19th century. Furthermore, we compared the performance of Resnet50 and Unet architectures. Pretrained Unet architecture achieves better performance when compared to the pretrained Resnet50 architecture for each road type. Classification accuracy is also affected by the representation of road legends. As mentioned before, the thickness, color or style of the road type legends may increase the difficulty of recognizing the road types. Therefore, for the challenging road types, the F_measure may decrease to 0.10. However, these results are comparable to the results seen in the literature for processing historical maps. These results show that our dataset is noisy. In future work, we plan to apply preprocessing techniques and detect and remove the names of the places to improve the performance of our system. We further plan to remove the background (which has varying colors depending on the land types) and contour lines that complicate the road detection process. We also plan to apply data augmentation and deep transfer learning to overcome the limited data problem of some road types in our future work.

REFERENCES

[1] (2020). *A Discovery Tool for Geospatial (GIS) Data*. [Online]. Available: https://library.stanford.edu/search-services/earthworks

[2] J. H. Uhl, S. Leyk, Y.-Y. Chiang, W. Duan, and C. A. Knoblock, "Automated extraction of human settlement patterns from historical topographic map series using weakly supervised convolutional neural networks," *IEEE Access*, vol. 8, pp. 6978–6996, 2020.

[3] J. H. Uhl, S. Leyk, Y.-Y. Chiang, W. Duan, and C. A. Knoblock, "Extracting human settlement footprint from historical topographic map series using context-based machine learning," in *Proc. 8th Int. Conf. Pattern Recognit. Syst. (ICPRS)*, 2017, pp. 1–6.

[4] Y.-Y. Chiang, W. Duan, S. Leyk, J. H. Uhl, and C. A. Knoblock, "Training deep learning models for geographic feature recognition from historical maps," in *Using Historical Maps in Scientific Studies*. Berlin, Germany: Springer, 2020, pp. 65–98.

[5] M. Saeedimoghaddam and T. F. Stepinski, "Automatic extraction of road intersection points from USGS historical map series using deep convolutional neural networks," *Int. J. Geographical Inf. Sci.*, vol. 34, no. 5, pp. 947–968, May 2020.

[6] P.-A. Herrault, D. Sheeren, M. Fauvel, and M. Paegelow, "Automatic extraction of forests from historical maps based on unsupervised classification in the cielab color space," in *Geographic Information Science at the Heart of Europe*. Berlin, Germany: Springer, 2013, pp. 95–112.

[7] U. Gimmi, C. Ginzler, M. Müller, and A. Psomas, "Assessing accuracy of forest cover information on historical maps," *Prace Geograficzne*, vol. 1, no. 146, pp. 7–18, 2016.

[8] M. M. Najafabadi, F. Villanustre, T. M. Khoshgoftaar, N. Seliya, R. Wald, and E. Muharemagic, "Deep learning applications and challenges in big data analytics," *J. Big Data*, vol. 2, no. 1, p. 1, Dec. 2015.

[9] G. Timár, S. Biszak, B. Székely, and G. Molnár, "Digitized maps of the Habsburg military surveys–overview of the project of arcanum ltd.(Hungary)," in *Preservation Digital Cartography*. Berlin, Germany: Springer, 2010, pp. 273–283.

[10] J.-L. Arnaud, "The long life of a 1: 200,000 map of central Europe and the Balkans," in *History Cartography*. Berlin, Germany: Springer, 2014, pp. 111–128.

[11] W. Göderle and M. Pfaffenthaler, "Spaces, times and actors in the 19th and 20th centuries," in *Dynamiken der Wissensproduktion*. Bielefeld, Germany: Transcript-Verlag, 2018, pp. 9–22.

[12] M. E. Kabadayı, P. Gerrits, O. Özkan, and T. Koak, "A preliminary attempt to construct a geospatial, multimodal Ottoman transport network for 1899," in *Spatial Webs: Mapping Anatolian Pasts for Research and the Public*. Istanbul, Turkey: Koç Univ. Press, 2021, pp. 115–128.

[13] *3rd Military Mapping Survey of Austria-Hungary*, Eötvös Loránd Univ., Budapest, Germany, 2021. [Online]. Available: http://lazarus.elte.hu/hun/digkonyv/topo/3felmeres.htm, 2021

[14] (2021). *Generalkarte of Mitteleuropa (GKM 1:200.000 der Franzisco-Josephinischen Landesaufnahme (3). Österreichische Landesaufnahme)*. [Online]. Available: https://commons.wikimedia.org/wiki/File:Abk%C3%_GKM.jpg

[15] W. Duan, Y.-Y. Chiang, C. A. Knoblock, V. Jain, D. Feldman, J. H. Uhl, and S. Leyk, "Automatic alignment of geographic features in contemporary vector data and historical maps," in *Proc. 1st Workshop Artif. Intell. Deep Learn. Geographic Knowl. Discovery*, 2017, pp. 45–54.

[16] W. Duan, Y.-Y. Chiang, S. Leyk, J. H. Uhl, and C. A. Knoblock, "Automatic alignment of contemporary vector data and georeferenced historical maps using reinforcement learning," *Int. J. Geographical Inf. Sci.*, vol. 34, no. 4, pp. 824–849, Apr. 2020.

[17] B. Budig and T. C. Van Dijk, "Active learning for classifying template matches in historical maps," in *Proc. Int. Conf. Discovery Sci.* Berlin, Germany: Springer, 2015, pp. 33–47.

[18] B. Budig, T. C. van Dijk, F. Feitsch, and M. G. Arteaga, "Polygon consensus: Smart crowdsourcing for extracting building footprints from historical maps," in *Proc. 24th ACM SIGSPATIAL Int. Conf. Adv. Geographic Inf. Syst.*, Oct. 2016, pp. 1–4.

[19] Esri. (2020). *Arcgis Pro (Version 2.5)*. [Online]. Available: https://www.esri.com/en-us/arcgis/products/arcgis-pro/overview

[20] *GDAL/OGR Geospatial Data Abstraction software Library*, Open Source Geospatial Foundation, Beaverton, OG, USA, 2021. [Online]. Available: https://gdal.org

[21] (2019). *Google, Tensorflow*. [Online]. Available: https://www.tensorflow.org/

[22] J. Ugarte Macías, "Deep learning: Segmentation of documents from the Archivo General de Indias with dhSegment and NeuralLineSegmenter," M.S. thesis, Dept. Signal Theory Commun., Univ. Sevilla, Seville, Spain, 2019.

[23] S. Ares Oliveira, B. Seguin, and F. Kaplan, "DhSegment: A generic deep-learning approach for document segmentation," in *Proc. 16th Int. Conf. Frontiers Handwriting Recognit. (ICFHR)*, Aug. 2018, pp. 7–12.

[24] Y. S. Can and M. E. Kabadayı, "CNN-based page segmentation and object classification for counting population in Ottoman archival documentation," *J. Imag.*, vol. 6, no. 5, pp. 32–49, 2020.

[25] Y. S. Can and M. E. Kabadayı, "Automatic cnn-based Arabic numeral spotting and handwritten digit recognition by using deep transfer learning in Ottoman population registers," *Appl. Sci.*, vol. 10, no. 16, pp. 5430–5446, 2020.

[26] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Assist. Intervent.* Berlin, Germany: Springer, 2015, pp. 234–241.

[27] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[28] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proc. 13th Int. Conf. Artif. Intell. Statist.*, 2010, pp. 249–256.

[29] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: http://arxiv.org/abs/1412.6980
[30] S. Ioffe, "Batch renormalization: Towards reducing minibatch dependence in batch-normalized models," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 1945–1953.
[31] K. V. Jobin and C. V. Jawahar, "Document image segmentation using deep features," in *Computer Vision, Pattern Recognition, Image Processing, and Graphics*, R. Rameshan, C. Arora, and S. Dutta Roy, Eds. Singapore: Springer, 2018, pp. 372–382.

YEKTA SAID CAN received the B.Sc. and M.Sc. degrees in computer engineering from the Faculty of Engineering, Boğaziçi University, Istanbul, Turkey, in 2012 and 2014, respectively, and the Ph.D. degree from the Department of Computer Engineering, Bogaziçi University, in 2020. For two years, he worked as a Researcher with the TUBITAK BILGEM. For six years, he has also worked as a Teaching Assistant with Boğaziçi University. He is currently working on retrieving information from Ottoman population registers by applying computer vision methods with Koç University, as a Postdoctoral Researcher. His research interests include watermarking, speech and speaker recognition, image segmentation, physiological signal processing, and machine learning.

PETRUS JOHANNES GERRITS received the bachelor's degree in (landscape-)archaeology from Vrije Universiteit Amsterdam, with a focus on GIS, and the M.A. degree with Koç University, in 2018. He is currently pursuing the Ph.D. degree in geospatial data science with the University of Glasgow, with a focus on the use of historical maps and census data to gain better understanding of the long-term role of transport in urbanization. During his master's degree, he was working with the Barcın Höyük Excavation and UrbanOccupationsOETR projects, where he gained an interest in spatial databases and computational analysis in archaeology and humanities.

M. ERDEM KABADAYI received the B.Sc. degree from Middle East Technical University, Ankara, the M.Sc. degree in economics from the University of Vienna, and the Ph.D. degree from the University of Munich, in 2008. Until 2015, he has mainly worked on the economic, financial, and labor history of the Ottoman Empire. Within the UrbanOccupationsOETR, he is pursuing his academic career further as an economic historian, within the fields of digital and geospatial humanities, focusing on the Ottoman Empire, Bulgaria, Greece, and Turkey.