Towards Historical Map Analysis using Deep Learning Techniques

Ladislav Lenc[1,2], Josef Baloun[1,2], Jiří Martínek[1,2], Pavel Král[1,2]

[1] Dept. of Computer Science & Engineering
Faculty of Applied Sciences
University of West Bohemia
Plzeň, Czech Republic
[2] NTIS - New Technologies for the Information Society
Faculty of Applied Sciences
University of West Bohemia
Plzeň, Czech Republic
{llenc,balounj,jimar,pkral}@kiv.zcu.cz

Abstract. This paper presents methods for automatic analysis of historical cadastral maps. The methods are developed as a part of a complex system for map digitisation, analysis and processing. Our goal is to detect important features in individual map sheets to allow their further processing and connecting the sheets into one seamless map that can be better presented online. We concentrate on detection of the map frame, which defines the important segment of the map sheet. Other crucial features are so-called inches that define the measuring scale of the map. We also detect the actual map area.
We assume that standard computer vision methods can improve results of deep learning methods. Therefore, we propose novel segmentation approaches that combine standard computer vision techniques with neural nets (NNs). For all the above-mentioned tasks, we evaluate and compare our so-called "Combined methods" with state-of-the-art methods based solely on neural networks. We have shown that combining the standard computer vision techniques with NNs can outperform the state-of-the-art approaches in the scenario when only little training data is available.
We have also created a novel annotated dataset that is used for network training and evaluation. This corpus is freely available for research purposes which represents another contribution of this paper.

Keywords: Historical Maps · Document Image Processing · CNN · Fully-convolutional Neural Networks

1 Introduction

Processing of historical documents is of a great interest. Various archival documents are stored in vast collections and are gradually digitised. The digitisation is the first step in the process of making the documents accessible and easily exploitable.

We focus on the processing of cadastral maps from the first half of the nineteenth century. The maps cover parts of the former Austria-Hungary empire – a region currently linked to the Czech Republic. The maps are hand-drawn and are available as a set of individual map sheets. The final goal of the map analysis and processing is to provide a seamless connection of the individual sheets.

The main goal is twofold. The first task is to construct a virtual grid of the map sheet positions to learn which ones are next to each other according to the historical coordinate system. The second goal consists in precise detection of map borders within the map sheets that belong to the same grid cell (map sheets on the border of cadastre areas). Thanks to the detected border, the neighbouring sheets can be seamlessly connected. The outside of the map is masked within this procedure. The final result of the task demonstrated on map sheets covering Pilsen city area is depicted in Figure 1.

Fig. 1. Example of a composed map of the city of Pilsen constructed from the individual map sheets (`https://archivnimapy.cuzk.cz/uazk/pohledy/archiv.html`)

The information about the position within the map coordinate system (in the grid) is provided by so-called nomenclature. The nomenclature is basically a label of the map sheet and it is usually located at the top right corner. It contains information about the map area name, relative position from the main meridian (west "W.C." or east "O.C."), and indices (marked by a Roman and Arabic numerals and a pair of letters $a - d$ and $e - i$) that uniquely identify the position in the coordinate system. For a more detailed overview of the projection and positioning, see [13].

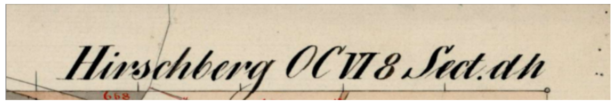

Fig. 2. Example of a nomenclature in Hirschberg cadastre area

A related task is detection of a map frame, which is ideally a rectangular area covering the grid cell. However, in reality, in the case of scanned hand-drawn maps, it is slightly curved which brings challenges for its precise detection. The measuring scale of the map is indicated by so-called inches that lie on the map frame and are equidistantly positioned.

This paper deals with several sub-tasks that allow us to achieve the seamless map connection. We concentrate mainly on detection of map frames and precise positions of inches and corners. We also solve map area segmentation. For all the above-mentioned tasks, we compare our so-called "Combined methods" with state-of-the-art methods based solely on neural networks. The proposed approaches combine standard computer vision techniques with neural networks (NNs). We assume that the combination of the standard computer vision techniques with NNs could outperform the state-of-the-art approaches in this particular scenario when only little training data are available.

2 Related Work

Map processing and analysis is usually comprised of segmentation and object detection algorithms. One important direction is the processing of historical maps. Several successful methods for map content segmentation were developed within the MapSeg competition [3]. The methods submitted to the competition are built on traditional computer vision approaches as well as neural network-based methods. IRISA [1] method is a representative of traditional methods while L3IRIS utilises a HSNet network [9]. The winning UWB method then combines both image processing methods and fully convolutional networks (FCN). One important aspect that the methods have in common is the adaptation on small amounts of training data which is common in this domain and hampers larger utilisation of deep learning methods. Lenc et al. [7] concentrated on the segmentation of the cadastre border and the detection of important points on it, so-called landmarks. The algorithms mainly relied on FCN networks and the results were post-processed and refined using image processing techniques such as mathematical morphology, skeletonization, etc. Chen et. al [4] presented a method for segmentation of historical maps. They utilised BiDirectional Cascade Network (BDCN) for detecting edges and at the same time filtering unwanted features such as text. Their dataset is a set of urban map sheets of a central area of Paris from the year 1925.

Another important related part of the map analysis research focuses on remote sensing. With the massive deployment of remote sensing technologies, the automatic processing of the images got a new dimension. Timilsina et al. [14] presented a method based on convolutional networks. It was tailored for detection of tree coverage of cadastre parcels. Another sub-task, road detection from aerial images, was addressed by Kestur et. al [6]. The authors used a novel U-shaped FCN (UFCN) model. Last but not least, Neyns and Canters [10] provided an overview of approaches to map urban vegetation from high-resolution remotely sensed data.

3 Dataset

The dataset used for our experiments comes out of the Map Border Dataset[3]. We have created new annotations for the solved tasks, namely for the inch and the corner detection and also for the map area segmentation. The inch and corner masks allow automatic generation of map frame ground-truths. We utilised the border annotations together with the inch annotations to semi-automatically create the map area ground-truths.

The dataset is split into training, testing and validation sections, which contain 69, 20 and 10 images, respectively and it is available as an extension of the above mentioned Map Border Dataset on the same website.

4 Corner Detection

The map frame detection is important for the connection of the neighbouring map sheets. As mentioned above, the map frame is a bounding rectangle that surrounds the map content. In reality, the shape is slightly distorted, which makes the task more challenging. An example of a map frame with inches (crop of the original map sheet) is shown in Figure 3.

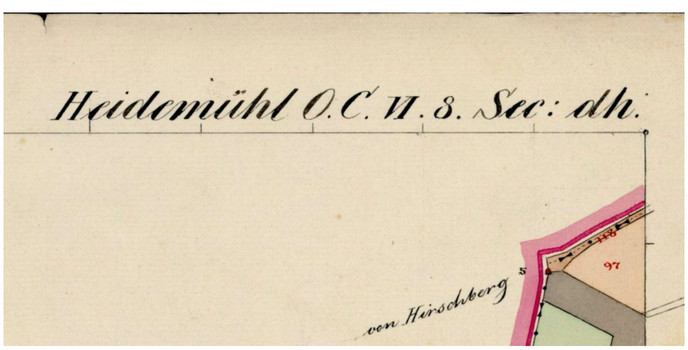

Fig. 3. Detail of a map frame with the inches

[3] `https://corpora.kiv.zcu.cz/map_border/`

A crucial step in the map frame detection is the detection of the map frame corners. The corner positions are essential for the frame representation and the subsequent tasks we have to perform. Namely, we utilize it to detect inches in the next step.

We proposed and implemented two methods for this task. The first one combines image processing techniques and an FCN while the second one relies solely on an FCN. We will report the approaches as "Combined" and "FCN-based".

4.1 Combined Method

We developed a novel approach that combines standard image processing techniques with neural networks. We first binarize the image using recursive Otsu thresholding method [11]. Next, we apply a fully convolutional network trained for prediction of map frame lines. We utilize the architecture proposed in [2]. The frame lines in the ground-truth masks are enhanced using a Gaussian kernel in order to put more weight to the line itself and gradually lower the weight with increasing distance from the line. This enhancement proved to be better compared to use solely the line in our preliminary experiments. The network prediction is then multiplied by the binarized image which leads to elimination of noise present in the binarization result.

After obtaining the rough map frame mask, the Hough transform is applied to detect horizontal and vertical lines. We filter all other lines with angles differing more than 2 degrees from the vertical and horizontal directions. This way we obtain several line candidates close to each of the map frame lines. Due to the fact that some map sheets may have more than two horizontal / vertical lines, we apply filtering based on the assumed distance of the line candidates. Thus we obtain only the real map frame lines. Intersections of the line candidates are chosen as corner candidates. As a final step, the candidate points are refined. We crop a rectangle surrounding the candidate corner point (see Figure 4) and construct horizontal and vertical projection profiles. The profile maxima are the coordinates of the actual corner positions.

Fig. 4. Binarized corner candidate region extracted around the candidate corner point

4.2 FCN-based Method

As a competing approach we have selected a method based solely on a neural network. We utilize the FCN network architecture proposed by Wick and

Puppe [15] The architecture is shown in Figure 5. This method is an adaptation of U-Net model proposed by Ronneberger et al. [12]. We chose this network because of its good results and more efficient computation times as shown in [8].

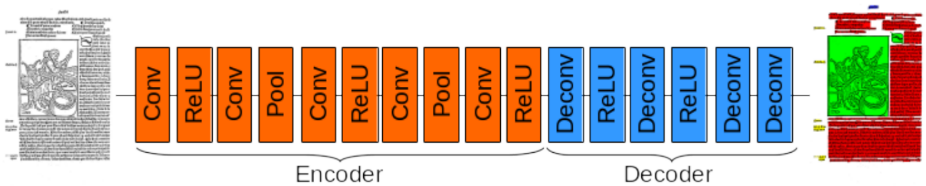

Fig. 5. Modified U-Net architecture proposed by Wick and Puppe [15]

Contrary to the U-Net architecture, it does not use skip connections. The whole architecture of this network is also simpler and the number of parameters is lower. The encoder part is composed of 5 convolutional and two pooling layers. The size of the convolution kernels is set to 5 and padding is used to keep the dimension. The decoder consists of 4 deconvolution layers.

The ground-truth masks for this task have circles at the positions of the corners. We thus predict the frame corner positions directly.

Due to the relatively small size of the ground-truth masks compared to the background, we utilize a patch-based approach that first divides the image into a set of rectangular patches according to a rectangular grid. The patches are then predicted individually and the final result is composed from the partial predictions.

4.3 Results

We have evaluated and compared both methods on the newly annotated map frame dataset. We report precision and recall of the detected corners as well as mean average error (MAE) measuring the average distance of the predicted corner and ground-truth position. Table 1 shows the results of this experiment.

The results indicate that the Combined method achieved better results regardless of the settings of the FCN-based method. The size of patches influences both precision and recall. The best results are obtained with patch sizes of 720×720 pixels. The number of patches per image used for training positively influences the overall results. An advantage of the FCN-based method is slightly better MAE which means that the predicted corners are closer to the ground-truth ones. This fact can be potentially used for refinement of the corner positions.

5 Inch Detection

The task of inch detection follows the corner detection. The inch positions are equidistantly placed on the map frame lines between the corners. In reality, the

Table 1. Comparison of corner detection results of the Combined method and the FCN-based method with different number of patches per image used for training and with different patch sizes

		P	R	MAE	IoU
Combined method					
		100	100	2.06	-
FCN-based method					
	Patch Size ($w \times h$)	**P**	**R**	**MAE**	**IoU**
25 patches	240×240	8.1	64.6	1.65	8.4
	320×320	11.2	72.9	1.61	11.1
	480×480	35.4	18.8	0.53	13.9
	640×640	82.0	54.2	1.61	35.8
	720×720	46.9	54.2	1.59	27.3
50 patches	240×240	4.8	52.1	1.59	4.7
	320×320	21.9	87.5	1.53	21.4
	480×480	35.9	43.8	2.72	18.2
	640×640	29.6	47.9	1.52	21.4
	720×720	64.7	81.3	1.57	46.4
100 patches	240×240	29.4	79.2	1.53	24.7
	320×320	33.3	77.1	1.45	28.4
	480×480	12.9	72.9	1.72	11.2
	640×640	69.2	72.9	1.33	43.3
	720×720	78.6	54.2	1.66	36.3
150 patches	240×240	16.8	85.4	1.80	16.1
	320×320	8.7	72.9	1.55	8.6
	480×480	70.5	83.3	1.23	53.0
	640×640	64.7	56.3	1.65	33.7
	720×720	86.0	58.3	1.58	42.5

distances vary slightly because the maps were hand-drawn. However, the known approximate distance can serve as a hint mainly in cases where the inches are not well-marked. There are also issues with false inches because there are usually many lines that can be misinterpreted as inches. We again compare a Combined method with a solely FCN-based one.

5.1 Combined Method

This method uses the outputs from the corner detection step. We first extract rectangular areas along the map frame lines. The crops are then used as input for an FCN trained for inch prediction. The network output is utilized for obtaining an initial set of inch candidates. Next we perform a check of the inches and try to compute positions of possibly missing ones. Once we have a complete set of candidate inches, a further refinement is applied. The procedure is very similar to the one used for corner position refinement. We again rely on the projection profiles computed in a small neighbourhood of the candidate position. Maxima of the projection profiles are used as the final inch positions.

5.2 FCN-based Method

The method utilizes the same network architecture and patch-based processing as the one used for corner detection. We mainly wanted to evaluate if it is possible to use only the FCN network and apply it on the whole image. This approach would contribute to the simplification of the overall task and could reduce computational demands of some computationally intensive steps in the corner detection algorithm, mainly the Hough transform and other related computations. In this case, it could fully substitute the corner detection step.

As in the case of corners, we use ground-truth in the form of masks with white dots at the positions of inches.

5.3 Results

Table 2 summarises the results obtained for the inch detection task. As in related studies, we report precision, recall, mean absolute error (MAE) and intersection over union (IoU) values.

Table 2. Comparison of inch detection results of the Combined and FCN-based method with different number of patches per image used for training and with different patch sizes

		P	**R**	**MAE**	**IoU**
Combined method					
		97.4	97.6	1.98	-
FCN-based method	**Patch Size ($w \times h$)**	**P**	**R**	**MAE**	**IoU**
50 patches	320×320	0.0	0	0	0
	160×640	73.0	34.3	0.93	0.12
	100×1280	70.1	32.6	1.34	0.11
	160×1280	0	0	0	0
100 patches	320×320	69.0	36.3	0.97	
	160×640	77.2	57.2	0.95	0.16
	100×1280	74.5	63.6	1.08	0.17
	160×1280	71.7	60.4	1.13	0.16
150 patches	320×320	65.2	34.8	1.33	0.08
	160×640	0	0	0	0
	100×1280	73.8	39.4	1.08	0.09
	160×1280	72.2	54.6	1.07	0.16

In this experiment, the solely FCN-based method proved relatively inefficient. The precision and recall values are relatively low. In this case, the patch size does not play as important role as in the corner detection. Similarly as in corner detection, the results of the FCN bring slightly more precise location of the points. The comparison proves that it is beneficial to use the pre-computed corner positions and detect inches only near the map frame lines. However, on the other hand, there is room to further improve the localisation accuracy of the Combined method.

6 Map Area Segmentation

The map area segmentation is another crucial step in the pipeline of the seamless map connection. It allows to visualize only the relevant area without blank parts outside the cadastre borders. Moreover, it is useful for attaching the relevant complementary part. Therefore, high demands are placed on the localization accuracy of the contours.

Generally, the map area might consist of several fragments within the map frame. The number of fragments is not limited, but in most cases, the frames contain only one fragment.

6.1 Combined Method

The Combined method utilises an FCN trained for prediction of fragment contours and multiplication with binarized image similarly as in Section 4.1. Given the fragment contour, we further use euclidean distance transform to obtain the distance from contour for each pixel. Based on that distance, the watershed is used to close the contour and retrieve the fragment area mask. This process is illustrated in Figure 6.

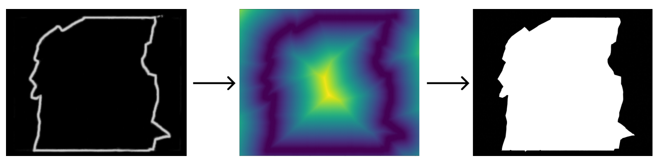

Fig. 6. Combined method process. From left: border contour, euclidean distance transform result, watershed result

Since several fragments can occur, we have to set watershed markers for each of them. These markers are decided as local maxima of the distance from the contour. That allows processing several fragments with precise details provided by binarization as illustrated in Fig. 7.

6.2 Mask R-CNN

Mask R-CNN [5] is a state-of-the-art model for image segmentation. It is basically a convolutional network trained for the detection of essential image regions. It is an extension of R-CNN and Faster R-CNN networks which predict solely bounding boxes of segmented objects. On the other hand, Mask R-CNN provides us also with segmentation masks that are necessary for our application. We utilise the implementation from the Detectron2 framework.

6.3 Results

Table 3 and Fig. 8 show the comparison of Mask R-CNN and our Combined method. The performance is measured in terms of intersection over union (IoU).

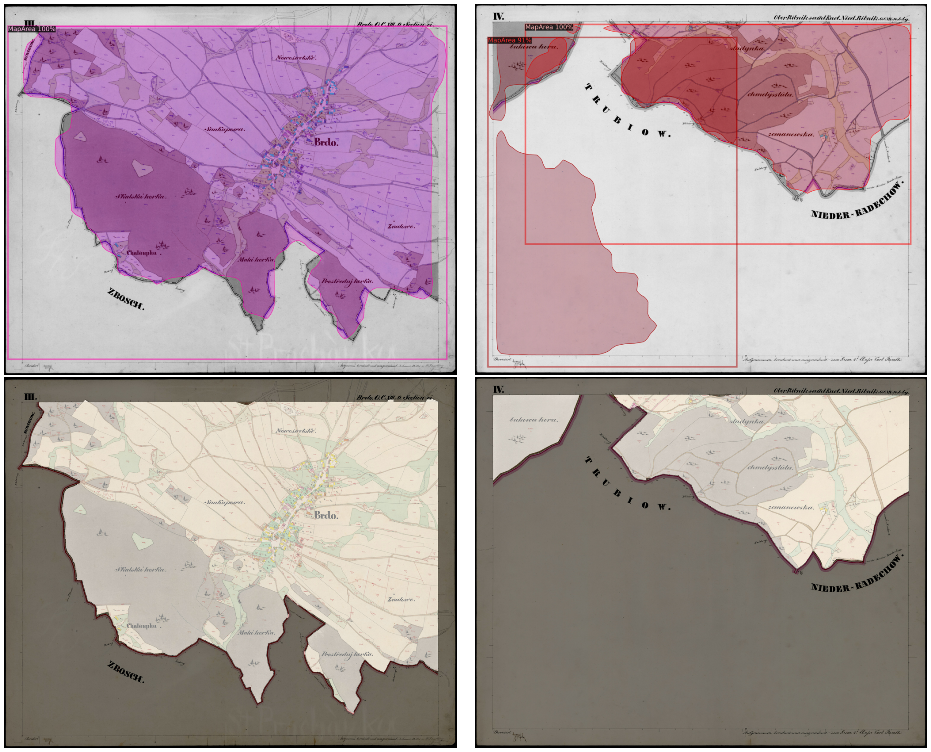

Fig. 7. Example of map area prediction using Mask R-CNN (top) and Combined method (bottom)

Table 3. Map area segmentation results of Combined method and Mask R-CNN

Method	IoU (%)	Hausdorff distance
Mask R-CNN	85.8	616.6
Combined method	97.0	172.9

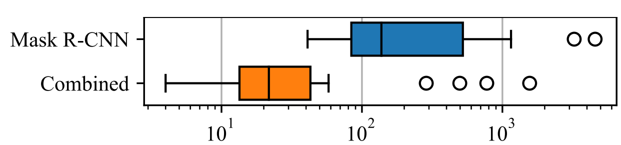

Fig. 8. Hausdorff distance distribution of test samples; smaller values are better

Further we provide Hausdorff distance (H) as error measure to compare the contour localisation accuracy. We can see that the Combined method significantly outperformed Mask R-CNN in this task.

Figure 7 shows two example predictions made by the Mask R-CNN network (top) and by the combined method (bottom). The visualizations confirm the results indicated by IoU and Hausdorff distance. We can observe that the Combined method performs much better mainly near the cadastre borders. This amount of detail is hard to achieve with Mask R-CNN or other FCN-based methods. The Combined method also has less false positive map regions. The differences are caused mainly by the different learning objectives when we predict only the border contours in the case of the Combined method. On the other hand, the Combined method may fail if the binarization or contour prediction is of poor quality. In such a case, a part or the whole fragment may be lost.

7 Conclusions and Future Work

We have presented a set of methods that will be utilized within a larger system for processing and analysis of historical cadastral maps. The final goal of the system is to seamlessly connect individual map sheets into one piece and allow a user-friendly presentation on the web. We have focused on the map frame identification which involves corner and inch marks detection. Another solved task is segmentation of map area where we need to differentiate between the useful map content and the outside area which we want to mask out.

We assumed that standard computer vision methods can improve the results of the deep learning approaches. Therefore, we proposed novel segmentation approaches that combine standard computer vision techniques with neural nets (NNs). For all the solved tasks, we evaluated and compared our so-called "Combined methods" with state-of-the-art methods based solely on neural networks.

The results have shown that it is beneficial to use a combination of neural networks with standard image processing techniques. We can state that utilizing the networks is important mainly in complicated cases where solely the traditional methods fail. On the other hand, using only deep learning is also not sufficient, mainly because of the need of large amounts of annotated data for network training. It is noticeable mainly in the task of map area segmentation, where the complicated state-of-the-art Mask R-CNN gives rather poor results compared to a simple FCN complemented by binarization, distance transform and watershed.

We have also created a novel annotated dataset, that is freely available for research purposes, which represents another contribution of this paper.

In the future, we would like to further improve the developed methods and thus improve the overall system for map connection. We also would like to experiment with networks based on the Transformer architecture, which lately brought interesting results also in the image processing domain.

Acknowledgement

This work has been partly supported by Grant No. SGS-2022-016 Advanced methods of data processing and analysis.

References

1. Aurelie, L., Jean, C.: Segmentation of historical maps without annotated data. In: The 6th International Workshop on Historical Document Imaging and Processing. pp. 19–24 (2021)
2. Baloun, J., Král, P., Lenc, L., Rocha, A., Steels, L., van den Herik, H.: Chronseg: Novel dataset for segmentation of handwritten historical chronicles. In: ICAART (2). pp. 314–322 (2021)
3. Chazalon, J., Carlinet, E., Chen, Y., Perret, J., Dumenieu, B., Mallet, C., Geraud, T., Nguyen, V., Nguyen, N., Baloun, J., Lenc, L., Kral, P.: ICDAR 2021 Competition on Historical Map Segmentation. In: Llados, J and Lopresti, D and Uchida, S (ed.) DOCUMENT ANALYSIS AND RECOGNITION, ICDAR 2021, PT IV. Lecture Notes in Computer Science, vol. 12824, pp. 693–707. IAPR (2021). https://doi.org/10.1007/978-3-030-86337-1_46, 16th IAPR International Conference on Document Analysis and Recognition (ICDAR), ELECTR NETWORK, SEP 05-10, 2021
4. Chen, Y., Carlinet, E., Chazalon, J., Mallet, C., Duménieu, B., Perret, J.: Combining deep learning and mathematical morphology for historical map segmentation. In: Discrete Geometry and Mathematical Morphology: First International Joint Conference, DGMM 2021, Uppsala, Sweden, May 24–27, 2021, Proceedings. pp. 79–92. Springer (2021)
5. He, K., Gkioxari, G., Dollár, P., Girshick, R.: Mask r-cnn. In: Proceedings of the IEEE international conference on computer vision. pp. 2961–2969 (2017)
6. Kestur, R., Farooq, S., Abdal, R., Mehraj, E., Narasipura, O.S., Mudigere, M.: Ufcn: A fully convolutional neural network for road extraction in rgb imagery acquired by remote sensing from an unmanned aerial vehicle. Journal of Applied Remote Sensing **12**(1), 016020 (2018)
7. Lenc, L., Prantl, M., Martínek, J., Král, P.: Border detection for seamless connection of historical cadastral maps. In: Document Analysis and Recognition–ICDAR 2021 Workshops: Lausanne, Switzerland, September 5–10, 2021, Proceedings, Part I 16. pp. 43–58. Springer (2021)
8. Martínek, J., Lenc, L., Král, P.: Building an efficient ocr system for historical documents with little training data. Neural Computing and Applications **32**, 17209–17227 (2020)
9. Min, J., Kang, D., Cho, M.: Hypercorrelation squeeze for few-shot segmentation. CoRR **abs/2104.01538** (2021), `https://arxiv.org/abs/2104.01538`
10. Neyns, R., Canters, F.: Mapping of urban vegetation with high-resolution remote sensing: A review. Remote sensing **14**(4), 1031 (2022)
11. Nina, O., Morse, B., Barrett, W.: A recursive otsu thresholding method for scanned document binarization. In: 2011 IEEE Workshop on Applications of Computer Vision (WACV). pp. 307–314. IEEE (2011)
12. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation. Miccai pp. 234–241 (2015)
13. Timár, G., Molnár, G., Székely, B., Biszak, S., Varga, J., Jankó, A.: Digitized maps of the Habsburg Empire The map sheets of the second military survey and their georeferenced version (01 2006). https://doi.org/10.13140/2.1.1423.7127
14. Timilsina, S., Sharma, S., Aryal, J.: Mapping urban trees within cadastral parcels using an object-based convolutional neural network. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences **4**, 111–117 (2019)

15. Wick, C., Puppe, F.: Fully convolutional neural networks for page segmentation of historical document images. In: 2018 13th IAPR International Workshop on Document Analysis Systems (DAS). pp. 287–292. IEEE (2018)