FCN-Boosted Historical Map Segmentation with Little Training Data

Josef Baloun[1,2][0000−0003−1923−5355], Ladislav Lenc[1,2][0000−0002−1066−7269], and Pavel Král[1,2][0000−0002−3096−675X]

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

Abstract. This paper deals with automatic image segmentation in poorly resourced areas. We concentrate on map content segmentation in historical maps as an example of such a domain. In such cases, conventional computer vision (CV) approaches fail in unexpected unique regions such as map content area exceeding the map frame, while deep learning methods lack boundary localization accuracy. Therefore, we propose an efficient approach that combines conventional CV techniques with deep learning and practically eliminates their drawbacks. To do so, we redefine the learning objective of a simple fully convolutional network to make the training easier and the model more robust even with few training samples. The presented method provides excellent results compared to more sophisticated but solely deep learning or traditional computer vision techniques as shown in "MapSeg" segmentation competition, where all other approaches were significantly outperformed. We further propose two additional approaches that improve the original method and set a new state-of-the-art result on the MapSeg dataset. The methods are further tested on an extended version of the Map Border dataset to show their robustness.

Keywords: Historical Map · Segmentation · Little Data.

1 Introduction

Historical maps owned by various national archives and libraries are a rich source of information. They often contain valuable and precisely plotted geographical entities. The digitized materials then offer a great potential for many historical studies [4] and they are beneficial for geographical information systems (GIS) communities for example. The process of map vectorization is thus of a great interest.

In the last decades, such maps have been gradually digitized and a lot of the materials are already accessible in an electronic form. However, the digitization is only the first step in the processing of the maps. There is a number of tasks

that have to be carried out in order to facilitate the search and fully utilize the materials which brings a number of research problems in the image processing field. The main emphasis is put on automatic approaches with good generalization abilities. However, this is problematic to achieve in some specific domains given the limited amount of annotated training data, which is almost always the case in the historical documents. It holds true in the case of historical maps as well. There are many differences between maps from different areas e.g. different colors, width of strokes, decorations and also map borders. Therefore, every collection is more or less unique. The manual annotation is time-consuming and therefore costly, which can explain also the lack of datasets in this domain.

The segmentation is an essential task which must be done after digitization. It allows further processing to focus only on the relevant area. Therefore, high demands are placed on the segmentation results. In this work, we concentrate on map content area detection according to the *Task 2* of the ICDAR 2021 Competition on Historical Map Segmentation [4] (*MapSeg*). The main goal is to provide a segmentation mask of a map content area (Fig. 1) and to remove features surrounding the actual map like map frame, legends and titles. Those elements are separated from the map content area by frames, however this is not the case for all of them. Moreover, the frames are frequently crooked or damaged and also exceeded by the map content area.

Although neural networks have remarkable learning capacity and achieved state-of-the-art results in many visual tasks including segmentation, their results still contain more or less errors in mask predictions given the low amount of training data. In the worst case, they do not work at all. On the other hand, it is also hard to handle document uniqueness with only conventional computer vision methods.

Therefore, we propose three methods with an easier learning objective that combines both conventional and deep learning approaches and lowers the effort needed for handcrafted features. Even though the methods utilize the simple FCN model, a significant improvement is achieved compared to much more sophisticated but purely deep learning or conventional computer vision techniques. We do not focus on different neural network architectures in this work, even though it may play a role. We instead focus on the combination of conventional and deep learning features in order to obtain the best results with a limited amount of data while minimizing the effort needed.

2 Related Works

Many methods were proposed for general image segmentation, e.g. fully convolutional networks (FCNs), where the input image and a segmentation mask are provided as a training sample. However, they rarely take into account the lack of data and the task specificity and only a few of them relate directly to maps. We first briefly report methods submitted to the MapSeg competition. Then we summarize methods solving map segmentation and analysis in general.

Fig. 1. MapSeg dataset sample: The map image overlayed with map content area ground-truth [5]

CMM [4] method is a representative of traditional approaches. The idea is to detect the contour lines and reconstruct them from the center of the image. It consists of several steps including the quasi-flat zone algorithm to eliminate map margin and a watershed to close the contour. *IRISA* [1] is another well performing traditional computer vision approach, that does not require any training. It relies on line segments. These segments are extracted from the image at various resolutions. Then, the grammatical rules are used to detect the map content area contour. *L3IRIS* [4] approach utilizes the state-of-the-art few-shot segmentation method HSNet [16].

The problem of map segmentation was solved for example in [14]. The authors proposed a method based on linear element features. A robust grid detection in historical maps relying on Hough transform is presented in [3]. The boom of neural networks and deep learning brought new possibilities for automatic segmentation and analysis of historical map resources. The potential usage of such methods was discussed in [10]. A method based on convolutional neural networks (*CNN*) was proposed in [15]. It uses an advanced guided watershed transform for obtaining superpixels [13]. A shallow CNN is then used for superpixel classification. A method for map segmentation utilizing handcrafted features, CNN and mathematical morphology was presented in [7]. A novel architecture for map segmentation was proposed in [9]. It has an encoder-decoder structure similarly as U-Net [19] and additionally uses cross-scale skip connec-

tions. A cadastre borders and important markers are detected in [12] utilizing an FCN or conventional computer vision techniques. A combination of deep learning and conventional computer vision methods is proposed in [8] to vectorize the historical maps. Names of cities and other landscape features are detected using several object detection models and further processed in [11].

3 Map Content Area Segmentation

Since deep learning can deal with hardly definable specialties in the map documents, we employ a simple model as a feature extractor to predict the borders of the area. We have identified experimentally that predicting only border contours is a much easier learning objective than predicting the whole map content area (Fig. 2). If we train the FCN model to predict the whole map content area, we want to predict every pixel there as positive (e.g. roads, buildings or text). There are also similar conflicting objects outside the map content that we want to predict as negative (legends for example). On the other hand, when predicting only border contours, the network can focus for example on lines, transitions between "empty" and "non-empty" areas, border decorations or legends. The number of possible input variants for a positive pixel is therefore much lower than in the previous case. We found this helpful to train the network.

The predicted border contour can be closed and transformed into the map content area utilizing morphological operations for example. At the same time, the conventional computer vision approaches can improve the results in terms of localization accuracy. Therefore, the three main steps of the methods are *border prediction*, *image binarization* to improve localization details and *post-processing* to close the contour.

First, we describe the border prediction and image binarization steps. Since the post-processing step changes across the approaches, details are provided within each specific approach.

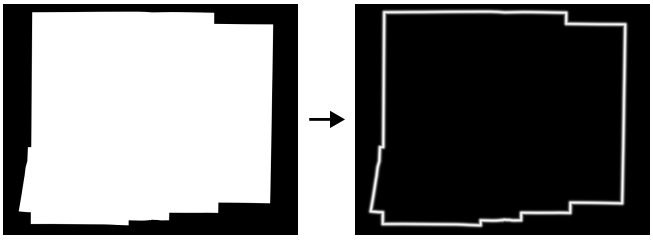

Fig. 2. Modified learning objective

3.1 Border Prediction

For the border prediction, we adapt a simplified U-Net-like FCN for general segmentation [2] as a feature extractor. As was shown in the paper, FCNs generalize well and they can also deal with the small amount of training samples. Furthermore, the border contours often appear close to image borders. That information can be utilized by the network thanks to the padding as discussed in the paper.

Compared to [2], the utilized network has half of the filters in the convolutional layers. The network's input is the whole down-sampled image and the output is the predicted mask of the borders. Since the input images are large, they are firstly eroded to propagate thin black lines and then down-sampled to fit 1024 px rectangle. The reason for the down-sampling is a compromise between network context capability, localization accuracy and computational costs. We refer to [2] for further details of the architecture.

The ground-truth was automatically generated from the provided map content area masks in the original image resolution (Fig. 2). The value for each pixel x was obtained using Gaussian function (Eq. 1), where $(x - b)$ stands for the distance to the border and σ is set to 50.

$$f(x) = exp\left(-\frac{1}{2}\frac{(x-b)^2}{\sigma^2}\right) \tag{1}$$

The reason for that is to provide more true positives and decrease the possibility of discontinuities in predictions. The uncertainty, that the pixel does not have to be strictly classified as the border or not, can also make classification more robust as discussed in [18]. We also use image augmentation techniques (mirroring, rotation and random distortion) to enlarge the training set.

3.2 Image Binarization

Since the borders are usually present in the input image, we found it useful to use them directly in order to have as precise results as possible. Therefore we adapt a recursive Otsu binarization method [17].

In a nutshell, the method firstly removes the background estimated by a median filter. It is ideal to propagate thin lines and also to discard large homogeneous areas. This step also allows the method to deal with the brightness inconsistency. After that, the image is recursively binarized using Otsu thresholding with hysteresis that reduces the amount of noise present in the binarized image. A drawback is a significant amount of remaining noise in the result making it difficult to process.

3.3 UWB Method

The winning method of the MapSeg competition is depicted in Fig. 3. The input image (Fig. 3.a) is binarized (Fig. 3.b) and the map content area borders (Fig. 3.d) are predicted in parallel. The prediction of map border is followed by post-processing resulting in an estimated mask of the map content area (Fig. 3.e).

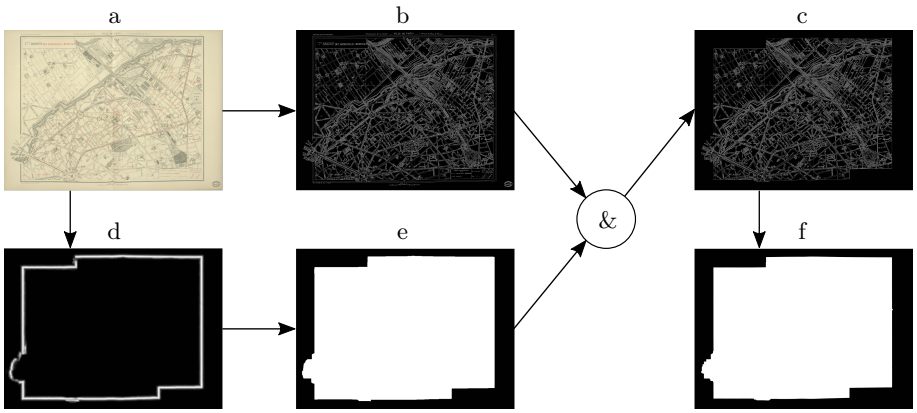

Fig. 3. Map content area segmentation process of UWB method: *(a)* input image, *(b)* binarized input image, *(c)* binarized image masked with estimated mask, *(d)* FCN border prediction, *(e)* estimated mask, *(f)* result

The estimated mask is then combined with binarized image utilizing logical and operation (Fig. 3.c). Finally, the post-processing is repeated to obtain map content area mask (Fig. 3.f).

The same post-processing is used for producing both masks Fig. 3.e and 3.f from the inputs Fig. 3.d and Fig. 3.c, respectively. It is similar to the morphological closing. It starts with dilation to fill eventual discontinuities. Then, the biggest connected component is selected and filled. Finally, the erosion is applied. The dilation and the erosion use the same rectangular kernel which is chosen with respect to the input images where map content borders usually follow horizontal and vertical lines.

A drawback of this method is the need to manually set the kernel. It is prone to improper settings and can cause fragmentation and errors if the contour is not properly closed. The post-processing also results in losing details as can be seen in Fig. 10. Therefore, we have proposed two improved versions of the method as described below.

3.4 BEW Method

To face the drawback of the baseline method, we use Border prediction, Euclidean distance transform and Watershed (*BEW*) according to Fig. 4. The method does not utilize any conventional features. Therefore, it can be used even if there is no possibility to extract the details conventionally.

A similar approach was presented in [8], but it fails if the contour is not properly closed. Therefore, we further extend the approach and use euclidean distance transform (Fig. 4.c). In that case, the missing fragments are fixed as illustrated in Fig. 5.

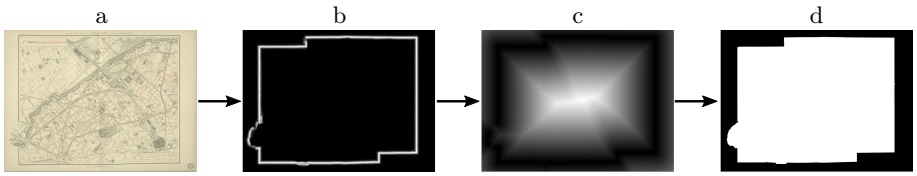

Fig. 4. Map content area segmentation process of BEW method: *(a)* input image, *(b)* FCN border prediction, *(c)* euclidean distance transform, *(d)* result using watershed

Fig. 5. From the left: detail of unclosed contour prediction, its euclidean distance transform, the result after watershed

3.5 BBEW Method

Optionally, the image binarization features can be used in order to improve localization accuracy as in the baseline method. The *BBEW* method uses Border prediction, Binarization, Euclidean distance transform and Watershed according to Fig. 6. The binarized image (Fig. 6.b) is masked with predicted borders

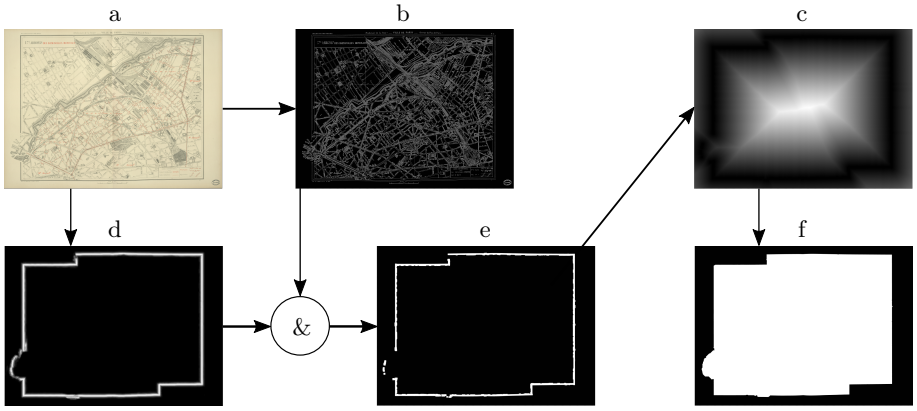

Fig. 6. Map content area segmentation process of BBEW method: *(a)* input image, *(b)* binarized input image, *(c)* euclidean distance transform, *(d)* FCN border prediction, *(e)* binarized image masked with predicted borders – dilated for visualization purposes, *(f)* result using watershed

(Fig. 6.d). Then, the euclidean distance transform is used to deal with unclosed contours. It provides a border distance matrix (Fig. 6.c). Finally, the watershed is used to segment the map content area (Fig. 6.f).

4 Experimental Set-up

In this section, we describe the dataset and evaluation criteria. For more details on the methods from Section 3, we refer to https://gitlab.kiv.zcu.cz/balounj/21_icdar_mapseg_competition, where the source codes and other related materials are freely available for non-commercial purposes.

4.1 MapSeg Competition Dataset

The map sheets constituting this dataset [5] are collected from 9 atlases of the City of Paris published between years 1894 and 1937. There are approximately 20 sheets for each year and the image resolution is very high (about 10000x10000 pixels).

The competition involved three tasks: Detect building blocks, Segment map content area and Locate graticule lines intersections. For each of these tasks, training, validation and test sets are available. For the second task, the sizes of train, validation and test sets are 26, 6 and 95 respectively. The dataset is available at https://zenodo.org/record/4817662.

4.2 Map Border Dataset

The Map Border dataset [12] consists of historical cadastral maps originating from the second half of the nineteenth century. It contains annotations for cadastre borders, important landmarks and other features for the task of border detection. We have extended the annotations with the map frame masks for the purposes of this work.

The image resolution is about 8400x6850 pixels. As illustrated in Fig. 7, the map sheets have different characteristics than the ones from MapSeg dataset.

We used 12 images for testing and 22 images for training and validation.

5 Evaluation criteria

For the reported results, we follow the MapSeg competition [4] scenario and also use the provided evaluation tools [6]. For the map content area segmentation evaluation, the 95th percentile variant of Hausdorff distance ($d_{H_{0.95}}$) is used as error measure. The final measure is the average of all test image measures.

We consider the Hausdorff distance appropriate for the task, since it focuses on shapes and details at the borders. The result is not distorted since it is not affected by the large area as in Intersection over Union for example.

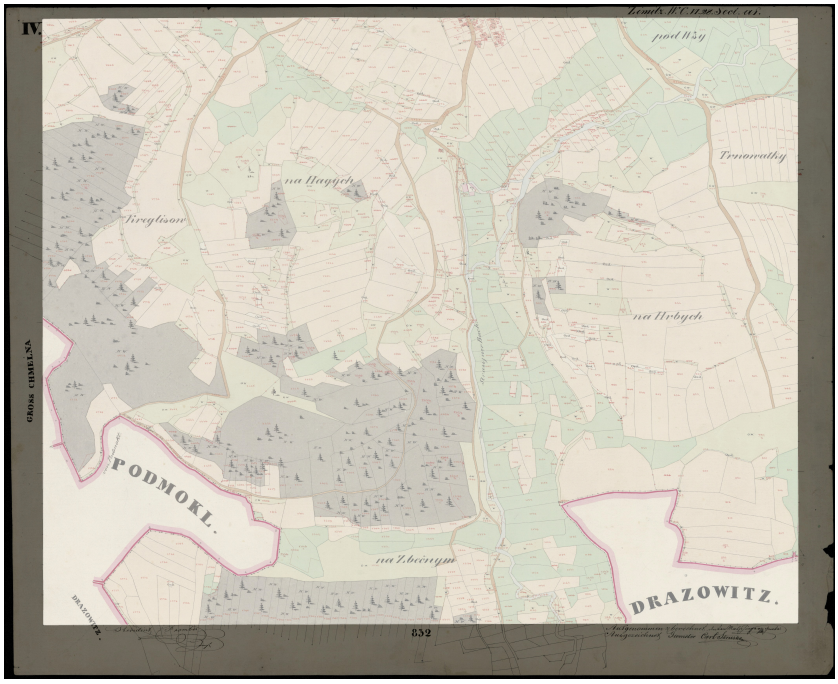

Fig. 7. Sample from Map Border dataset

6 Results

In this section, we report the obtained results and compare our *UWB*, *BEW* and *BBEW* methods with the best methods on the MapSeg dataset [5]. The proposed methods are further verified on the Map Border dataset.

As can be seen in Table 1, our methods surpassed the other methods by a significant margin. The proposed methods show excellent results even on the noisy historical map images.

Method	MapSeg dataset	Map Border dataset
CMM	85	–
IRISA	112	–
L3IRIS	126	–
UWB (Ours)	19.0	25.5
BEW (Ours)	18.1	12.0
BBEW (Ours)	**12.0**	**9.5**

Table 1. Final Hausdorff error ($d_{H_{0.95}}$ [px]) for map content area segmentation task

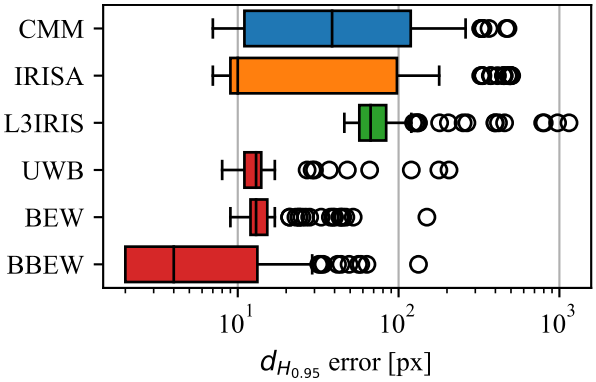

Fig. 8. Test images error distribution for map content area segmentation task on MapSeg dataset (Ours in red color)

The conventional approaches (*CMM* and *IRISA*) are very precise in the areas that contain the border contours, but they can hardly deal with other unique areas like map content exceeding the frame and some legends. This leads to higher variance in the errors in Fig. 8.

The deep learning approach (*L3IRIS*) has higher median value but smaller variance of errors. It usually catches the unique areas but it is missing the precision at the borders and the localization accuracy is not very convincing.

In the same figure, our approaches have small median value and also small variance of errors. They profit from both conventional and deep learning approaches and practically eliminate their drawbacks. On the other hand, the outliers in Fig. 8 are usually caused by wrongly predicted legends and are still present in each approach. This could be probably solved by extending the training part of the dataset, improved augmentation or further post-processing.

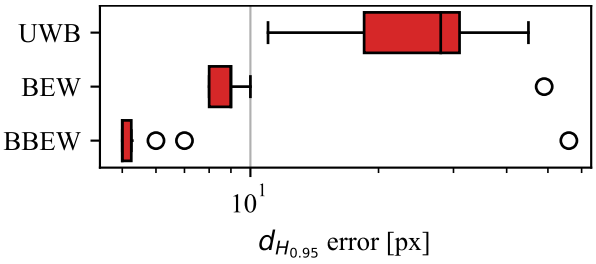

Fig. 9. Test images error distribution for map content area segmentation task on Map Border dataset

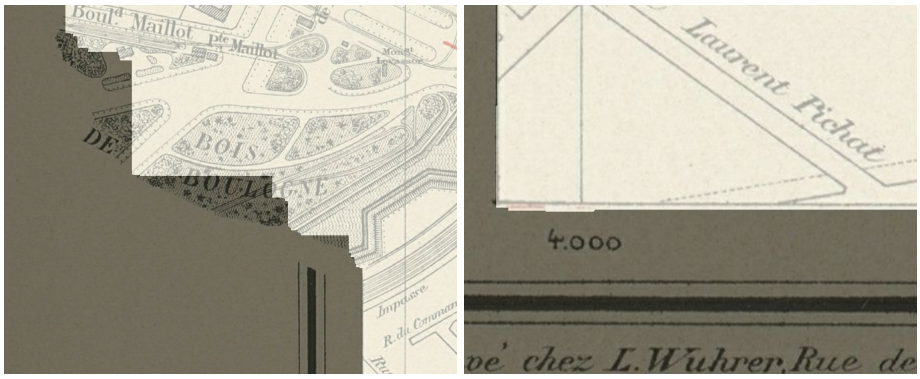

Fig. 10. Details of UWB method result

Fig. 11. Details of BEW method result

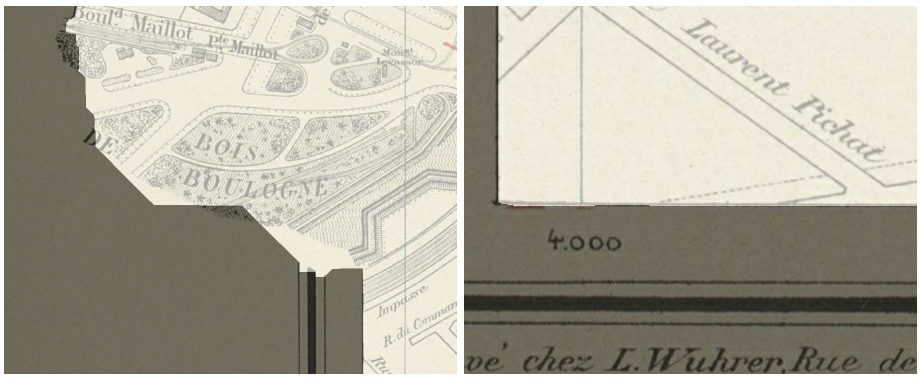

Fig. 12. Details of BBEW method result

The UWB method provides good results and allows to utilize the details provided by image binarization step. But naive post-processing has several drawbacks that can cause errors and loss of detail as in Fig. 10. As presented in Fig. 9, this issue is more evident on the results from Map Border dataset.

The BBEW method solves these drawbacks and provides as many details as possible while preventing the erroneous closing of the contour. It sets the new state-of-the-art result of $d_{H_{0.95}} = 12.0$. As illustrated in Fig. 12, the amount of detail is fascinating, especially in areas where border features can be obtained conventionally and further combined with deep learning features.

Interesting observation is a very good result of *BEW* method (Fig. 11). It indicates the advantage of the modified learning objective. *L3IRIS* utilizing the state-of-the-art few-shot segmentation method fails compared to *BEW* utilizing the simple general FCN segmentation model with modified learning objective and contour closing. It is obvious, that the border prediction is a much more suitable learning objective for the task and it has probably a bigger impact than the selection of the model.

7 Conclusions

In this paper, we faced the segmentation of historical maps. In such poorly resourced domain, common deep learning approaches often fail due to the lack of training data or even related data that could be used for transfer learning.

We have proposed three efficient map segmentation approaches that utilize an easier learning objective for a general FCN and post-processing to get the original objective. It allows them to work even with little training data while producing excellent results that can be optionally refined using conventional image binarization features.

The proposed methods are evaluated on the MapSeg dataset which was used in the ICDAR 2021 segmentation competition. We have shown that the proposed methods outperform significantly all other approaches and that the more suitable learning objective may have a bigger impact than the choice of a deep learning model. With *BBEW* method, we set the new state-of-the-art of $d_{H_{0.95}} = 12.0$ on MapSeg dataset. We further verified the methods on the Map Border dataset with corresponding results. Thus, the combination of deep learning with conventional computer vision techniques seems very promising, especially for poorly resourced domains such as historical documents.

Another contribution of our work consists in the availability of the source codes for the research purposes.

Acknowledgements This work has been partly supported by the 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.: Chronseg: Novel dataset for segmentation of handwritten historical chronicles. In: Proceedings of the 13th International Conference on Agents and Artificial Intelligence - Volume 2: ICAART,. pp. 314–322. INSTICC, SciTePress (2021). https://doi.org/10.5220/0010317203140322
3. Baloun, J., Lenc, L., KrÃąl, P.: Robust grid detection in historical map images. In: 2022 IEEE International Conference on Image Processing (ICIP). pp. 1931–1935 (2022). https://doi.org/10.1109/ICIP46576.2022.9897721
4. 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
5. Chazalon, J., Carlinet, E., Chen, Y., Perret, J., Mallet, C., Duménieu, B., Géraud, T.: Icdar 2021 competition on historical map segmentation – dataset. online dataset (2021). https://doi.org/10.5281/zenodo.4817662
6. Chazalon, J., Edwin-Carlinet: icdar21-mapseg/icdar21-mapseg-eval: Zenodo archival (2021). https://doi.org/10.5281/zenodo.4818400
7. Chen, Y., Carlinet, E., Chazalon, J., Mallet, C., Duménieu, B., Perret, J.: Combining deep learning and mathematical morphology for historical map segmentation. In: International Conference on Discrete Geometry and Mathematical Morphology. pp. 79–92. Springer (2021)
8. Chen, Y., Carlinet, E., Chazalon, J., Mallet, C., Dumenieu, B., Perret, J.: Vectorization of historical maps using deep edge filtering and closed shape extraction. In: Document Analysis and Recognition–ICDAR 2021: 16th International Conference, Lausanne, Switzerland, September 5–10, 2021, Proceedings, Part IV. pp. 510–525. Springer (2021)
9. Foroughi, F., Wang, J., Nemati, A., Chen, Z., Pei, H.: Mapsegnet: A fully automated model based on the encoder-decoder architecture for indoor map segmentation. IEEE Access (2021)
10. Garcia-Molsosa, A., Orengo, H.A., Lawrence, D., Philip, G., Hopper, K., Petrie, C.A.: Potential of deep learning segmentation for the extraction of archaeological features from historical map series. Archaeological Prospection **28**(2), 187–199 (2021)
11. Lenc, L., Martínek, J., Baloun, J., Prantl, M., Král, P.: Historical map toponym extraction for efficient information retrieval. In: Uchida, S., Barney, E., Eglin, V. (eds.) Document Analysis Systems. pp. 171–183. Springer International Publishing, Cham (2022)
12. Lenc, L., Prantl, M., Martínek, J., Král, P.: Border detection for seamless connection of historical cadastral maps. In: International Conference on Document Analysis and Recognition. pp. 43–58. Springer (2021)
13. Li, Z., Chen, J.: Superpixel segmentation using linear spectral clustering. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1356–1363 (2015)

14. Liu, T., Miao, Q., Xu, P., Song, J., Quan, Y.: Color topographical map segmentation algorithm based on linear element features. Multimedia Tools and Applications **75**(10), 5417–5438 (2016)
15. Liu, T., Miao, Q., Xu, P., Zhang, S.: Superpixel-based shallow convolutional neural network (sscnn) for scanned topographic map segmentation. Remote Sensing **12**(20), 3421 (2020)
16. Min, J., Kang, D., Cho, M.: Hypercorrelation squeeze for few-shot segmentation. CoRR **abs/2104.01538** (2021), https://arxiv.org/abs/2104.01538
17. 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)
18. Peterson, J.C., Battleday, R.M., Griffiths, T.L., Russakovsky, O.: Human uncertainty makes classification more robust. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) (October 2019)
19. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. pp. 234–241. Springer International Publishing, Cham (2015)