CNN-based Watershed Marker Extraction for Brick Segmentation in Masonry Walls

Yahya Ibrahim[1], Balázs Nagy[1,2], and Csaba Benedek[2,1]

[1] Péter Pázmány Catholic University, Faculty of Information Technology and Bionics
[2] Institute for Computer Science and Control, Hungarian Academy of Sciences

Abstract. Nowadays there is an increasing need for using artificial intelligence techniques in image-based documentation and survey in archeology, architecture or civil engineering applications. Brick segmentation is an important initial step in the documentation and analysis of masonry wall images. However, due to the heterogeneous material, size, shape and arrangement of the bricks, it is highly challenging to develop a widely adoptable solution for the problem via conventional geometric and radiometry based approaches. In this paper, we propose a new technique which combines the strength of deep learning for brick seed localization, and the Watershed algorithm for accurate instance segmentation. More specifically, we adopt a U-Net-based delineation algorithm for robust marker generation in the Watershed process, which provides as output the accurate contours of the individual bricks, and also separates them from the mortar regions. For training the network and evaluating our results, we created a new test dataset which consist of 162 hand-labeled images of various wall categories. Quantitative evaluation is provided both at instance and at pixel level, and the results are compared to two reference methods proposed for wall delineation, and to a morphology based brick segmentation approach. The experimental results showed the advantages of the proposed U-Net markered Watershed method, providing average F1-scores above 80%.

Keywords: Documentation application · Brick segmentation · Deep learning · U-Net · Watershed .

1 Introduction

Image-based analysis of man-built structures is considered as a core step of many applications, such as stability analysis in civil engineering, condition estimation and damage detection of buildings in architecture, digital documentation in archeology or maintenance and restoration in cultural heritage preservation. The surveyors in these processes need to extract comprehensive information about the studied sites, among others about the current conditions, the possible further actions, types of appropriate treatment, and the expected consequences of any intervention.

By investigating building masonry - either ancient walls or modern buildings - accurate detection and outlining of their structural components is a key initial step of the documentation process. The separation is based the fact that a masonry wall is a heterogeneous material that contains individual units (bricks, blocks, ashlars, irregular stones and others) and joints (mortar, clay, chalk etc.) between the main units, which bind the units together. Note that for simpler discussion, in our paper we use henceforward the term *brick* to describe the main units of any type, and the term *mortar* to describe the joints.

Manual segmentation of a large set of masonry images is time-consuming due the great number of bricks in an image, and such labeling is often inaccurate and unreproducible, as the output also depends on the experience of the operator. Thus, there is an increasing need nowadays for efficient automated processing. Moreover, due to various threats on the buildings (wars, natural disasters, etc.), quick and accurate documentation has became particularly necessary on precious archeology sites in the last decade.

In the paper we present a novel image based automated brick segmentation approach, which combines the strength of deep learning algorithms for robust brick localization under highly varying conditions, and the classical Watershed algorithm for accurate brick outline extraction and removal of mortar regions. We also introduce a new manually labeled dataset of diverse masonry images, which is used to train and test our approach, enabling its quantitative comparison to the state-of-the-art.

The paper is structured as follows. In Sec. 2 we summarize related work and the main contributions of the proposed approach. In Sec. 3 we introduce the proposed algorithm in details, while we present a detailed qualtitative and quantitative evaluation in Sec. 4. We conclude our work in Sec. 5, with mentioning remarks for future work.

2 Related work

In an earlier study [4] adopt various pixel-based and object-oriented image processing technologies for detecting and characterizing the structural damage in historical buildings based on multi-spectral measurements. Oses et. al. [8] focus on the classification of built heritage masonry for determining the necessary degree of protection in different buildings, using an automatic image-based delineation method. Riveiro et. al. [9] present an automatic color-based algorithm for segmenting masonry structures, based on an improved marker-controlled Watershed. However, this later algorithm [9] purely focuses on the morphological analysis of quasi-periodic masonry walls, where the geometry of masonry courses follows horizontal rows, which condition does not hold very often - especially for ancient walls (see Fig 4). Sithole et. al. [12] propose a semi-automatic segmentation algorithm to detect the bricks in masonry walls, working on *3D point cloud data* obtained by laser scanning. Although using such data sources becomes widespread in archeology and architecture nowadays, the 2D image based investigation addressed in this paper still has significance in particular for processing archive measurements, or in situations when long scanning surveys are not feasible. Since the method of [12] is based on the 3D triangulation of the 3D point cloud, reflectance, and RGB triplets, it cannot be suit to 2D data in a straightforward way. Similar issues appear by the work of Bosché et. al. [2], who introduce a method that simultaneously considers 3D information both in global and local levels for segmenting the walls into regions corresponding to bricks and mortar joints.

Focusing on the development of a widely applicable and purely 2D image based approach, automatic brick segmentation in masonry images is a notably challenging task due to several reasons [12] like similarity in surface texture between the bricks and the mortar, challenges caused by the lighting conditions, varying dimension and shape of the bricks and the mortar regions. Moreover, walls of different ages, built from various materials, and having different status appear significantly differently in the photos (see Fig. 2, 3(a)).

While deep learning (DL) algorithms have achieved remarkable success in various computer vision applications (segmentation, classification, detections, etc.) in a wide range of fields from medical images [14], scene understandings [13] or autonomous driving [1], we only find a few references yet for application of DL methods in architecture or cultural heritage documentation. In addition, existing methods [3,6] use deep learning rather for classification of the available images of the architectural heritage, instead of segmentation and feature extraction for detailed analysis.

The widely used Watershed algorithm [10] is a classical mathematical morphological method for image segmentation, and instance separation of definite object classes. The recently proposed Deep Watershed (DW) technique [1] realizes a possible way of combining deep learning with Watershed, however that method differs from our solution both from a methodological point of view, and in terms of the desired output. DW expects as input besides the raw camera image an accurate region mask obtained by semantic segmentation, which contains several touching or partially occluded instances of a given object class, such as vehicles or pedestrians in a traffic scene. Therefore, DW aims to particionate each homogeneous blob of the semantic map into tightly connected individual objects, which is implemented through a cascade usage of a direction network for estimating a vector from each object candidate point to the respective object center, and a second net which aims to approximate the Watershed energy for instance separation. On the other hand, in our case the brick instances are usually *not* tightly connected since they are mostly separated by the mortar regions, thus brick detection within a wall segment includes the separation the individual bricks from the mortar and from the neighboring bricks, which are sometimes slightly contacting (see Fig. 5). Another difference is that instead of training different Convolutional Neural Networks (CNNs) for region level semantic segmentation of the input image, and instance separation within the regions of interest (ROIs) [1], we use a single CNN that simultaneously provides information for ROI filtering, mortar removal and brick separation. Details of the proposed method are presented in the next section.

3 Proposed Method

The goal of the proposed method is to automatically extract the individual brick instances from 2D images taken from masonry walls of any types. Our approach is based on the widely used Watershed image segmentation algorithm [10], which considers the image as a topographic surface, where intensities of pixels correspond to altitude values. In the classical form of technique [10], local minima are extracted in the altitude map, then watersheds are defined by the lines that separate adjacent minima basins. As a usual artifact of the approach, the Watershed transform itself yields to oversegment masonry images, since the color values may strongly vary within the individual brick and mortar regions. A straightforward extension is applying markers in the process [10], so that we only enable flooding from specific seed points or internal markers taken inside each individual (brick) object, meantime we also use external markers for the mortar regions, which do not correspond to any brick instances.

The proposed algorithm can be divided into two main steps. The first one is a CNN-based delineation step, which aims to separate the regions of bricks from the mortal and other background regions in the image. The second step implements the Watershed-based segmentation of brick instances using external and internal markers extracted form the delineation output.

For giving a complete overview, Fig. 1 shows the dataflow of our method, with demonstrating the results of the subsequent filtering steps for a selected input image. The algorithms marked with arrows are detailed in Sections 3.1, 3.2.

3.1 Deep Learning Based Delineation for Separation of Brick and Mortar Regions

The first step of our approach is the delineation of the bricks in order to distinguish them from the mortar. Applying conventional edge detectors (such as Sobel or Canny) to tackle the problem is a straightforward solution in the literature, however experiments show that their performance is highly sensitive to image noise and contrast parameters. Riveiro's delineation algorithm [9] relies on the gradient of the pixel intensity attribute for making this distinction step. Oses et. al. [8] have proposed a heuristic algorithm that describes the geometric arrangement of blocks in the wall by a set of straight segments, which are extracted from image region boundaries that are created by using histogram based image quantization. However, even by using these complex delineation algorithms, the efficiency is limited to some specific (regular) wall structures [9], or the output is only used for classification of various sorts of masonry walls instead of segmenting the structural elements [8]. (We present later comparative tests between our algorithm and these methods in Sec. 4.3.)

Instead of using noisy hand-crafted features, our key idea is to solve the delineation task by a Convolutional Neural Network (CNN). Our choice was the U-Net network proposed by Ronneberger et al [11] for biomedical image segmentation, however it has already been applied with a great success in many segmentation tasks, and in addition it is able to reach high accuracy using a relatively small dataset. The U-Net architecture (see Fig. 1 (a)) can be decomposed into two main parts: (i) the encoder part consists of several convolution layers with ReLU (Rectified Linear Unit) activation function followed by max pooling layers while the decoder part (ii) consists of up-convolution and convolution layers. (i) encodes the image into a compact feature representation then (ii) decodes the features into a gray scale image which represents the local probabilities of the predicted classes.

To train the network we need input-target pairs. In our case the input is a 3-channel RGB image (raw photo), while the target is a binary mask where white pixels correspond to the brick regions, and black ones to the mortar between the bricks and to the background of the masonry walls (see Fig. 2). The prediction of the U-Net is a grayscale image, where high intensity values indicate the brick regions (Fig. 3(d)).

From a technical point of view, it should be noted that we trained U-Net using the Adam optimizer (Adaptive moment Estimation) with a binary cross entropy cost function and the number of epochs was 100.

3.2 Brick Instance Segmentation By a Sketch-driven Watershed Algorithm

The second step of the proposed algorithm is responsible for extracting the accurate outlines of the brick instances, relying on the previously obtained grayscale delineation map. As the U-Net outputs in Fig. 3(d) and 4(b) demonstrate, the delineation maps are quite reliable, however they might be noisy near to the brick boundaries, and some bricks are contacting, making simple connected component analysis (CCA) based separation prone to errors. To overcome these drawbacks, we apply a marker based Watershed [10] segmentation. (In Sec. 4.3 we also illustrate experimentally the difference between using CCA and Watershed algorithm.)

First, we binarize the delineation map via simple thresholding, and calculate the *inverted distance transform* (IDT) map of the obtained binary mask (M). Our aim is to extract a single compact seed region within each brick instance, which can be used as internal marker for the Watershed algorithm. Since the IDT map may have several false local minima, we apply the H-minima transform [7], which suppresses all minima under a given H-value (used $H = 5$ pixels). The H–minima supression step is illustrated in 1D in Fig. 1(b) and (c).

Finally, we apply flooding from the obtained H-minima regions, so that we consider the inverse of M (i.e. all mortar or non-wall pixels) as an external marker map, whose pixels cannot be assigned to any bricks. Results of the obtained brick contours are displayed over the input images in Fig. 4(c).

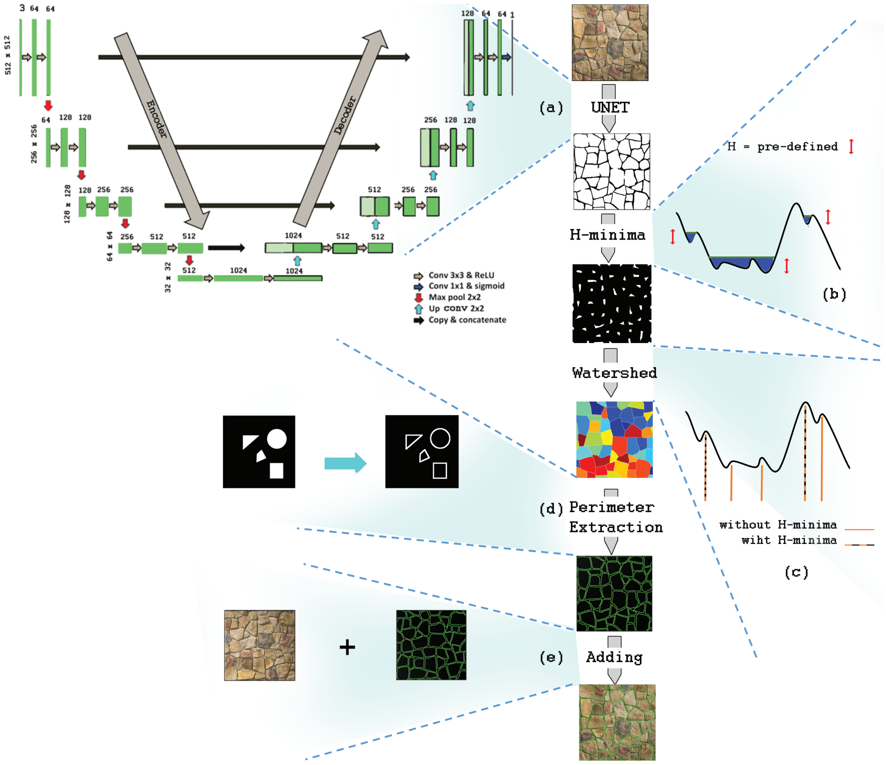
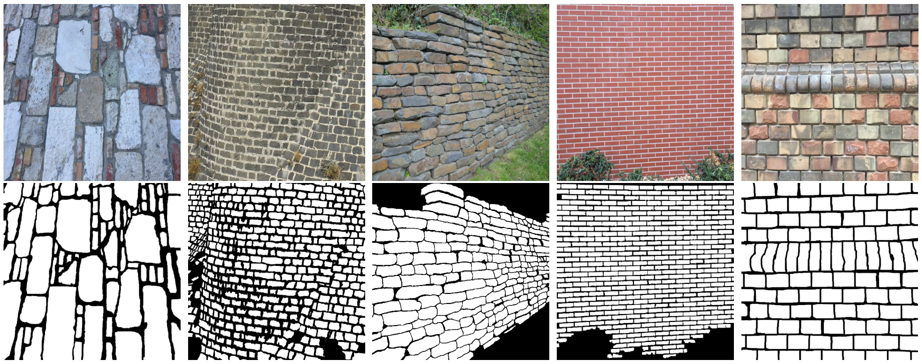


Fig. 1: Dataflow of our method. (a) U-Net model (b) H-minima algorithm (c) The Watershed H-minima (d) Perimeter Extraction (e) Adding step.



(a) Random (b) Square (c) Dry (d) Fine (e) Rough

Fig. 2: Samples from the dataset and with displaying the ground truth labels, (a) Random rubble masonry, (b) Square rubble masonry, (c) Dry rubble masonry, (d) Ashlar fine, (e) Ashlar rough.

4 Experiment and Results

4.1 Dataset Generation

For training and evaluation of the proposed U-Net markered Watershed technique, we have created a new annotated dataset containing images of masonry walls from various locations, including both ancient walls and facades of new-fashioned modern buildings. Depending on the wall structures, we divided the images into five main classes: three types of rubble masonry (Random, Square, Dry), and two types of ashlar (Fine, Rough). Note that the samples within each class may still differ in shape and size parameters of the bricks, while the mortar regions can be thin, tick or completely missing. Meantime, in the images the walls are visible from different viewpoints, while different lighting conditions and shadowing effects can be present. The new dataset contains overall 162 manually labeled images of size 512×512, where we used 117 images for training, 13 for validation and 32 for testing the algorithm.

Since large training set generation is a key issue for deep learning based methods, we also applied data augmentation on the training set by randomly rotating the images up to $+90°$, randomly shifting the images in horizontal and vertical direction with an offset up to 5% of the image width, and optionally applied shearing and horizontal flipping transforms, and zooming up to 5% of the size of the image. We have also augmented the test image set in a similar manner to expand the relevance of evaluation. The annotation follows the way described in Sec. 3.1. The images are labeled with two classes: background (black pixels in the mask) represents the mortar and the non-wall regions in the images, and foreground (white) represents the bricks. For each wall class, a sample image with the corresponding annotated mask is shown in Fig. 2.

4.2 Evaluation Methodology

In the experimental phase, we separately analyze the efficiency of the delineation step, and the final brick segmentation results.

To evaluate the delineation, we compare pixel-wise the output map of our U-Net component and other state-of-the-art methods (see Fig.3(b)-(d)) to the expected Ground Truth masks (Fig.3(e)). Since the correctness of the structure can be better described by the pixel-level accuracy of the thinner mortar regions, we calculate pixel-level Precision (Pr) and Recall (Rc) values from the viewpoints of the mortar pixels, and take the F1-score as the harmonic mean of Pr and Rc.

For evaluating the final brick segmentation results by the markered Watershed algorithm, we calculate both (i) object level and (ii) pixel level metrics. First of all, an unambiguous assignment is taken between the detected bricks (DB) and the ground truth (GT) bricks (i.e every GT object is matched to at most one DB candidate). To find the optimal assignment we use the Hungarian algorithm [5], where for a given DB and GT object pair the quality of matching is proportional to the intersection of union (IOU) between them, and we only take into account the pairs that have IOU higher then a pre-defined threshold 0.5. Therafter, object and pixel level matching rates are calculated as follows:

(i) At object (brick) level, we count the number of True Positive (TP), False Positive (FP) and False Negative (FN) hits, and compute the *object level* precision, recall and F1-score values. Here TP corresponds to the number detected bricks (DBs) which are correctly matched to the corresponding GT objects, FP refers to DBs which do not have GT pair, while FN includes GT objects without any matches among the DBs.

(ii) At pixel level, for each correctly matched DB-GT object pair, we consider the pixels of their intersection as True Positive (TP) hits, the pixels that are in the predicted brick but not in the GT as False Positive (FP), and pixels of the GT object missing from the DB as False Negative (FN). Thereafter we compute the *pixel level* evaluation metrics precision, recall and F1-score and the intersection of unions (IOU). Finally the evaluation metric values for the individual objects are averaged over all bricks, by weighting each brick with its total area.

4.3 Performance Evaluation

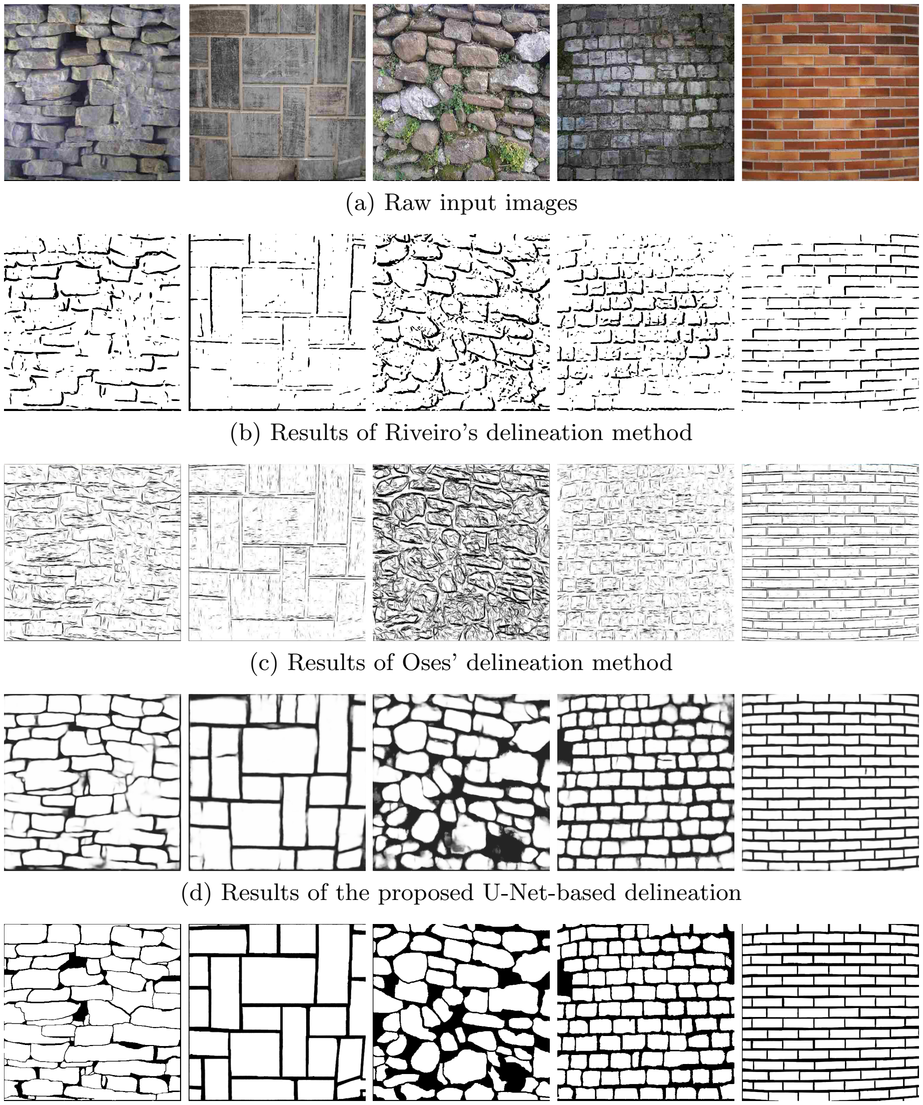
In this section, we analyze the performance of the proposed approach on the test data, based on the different evaluation parameters defined in Sec. 4.2.

We start with the discussion of the delineation step. Demonstrative sample results of using the state-of-the-art delineation methods and our method are shown in Fig. 3, and the corresponding quantitative evaluation values are provided in Table. 1. We can confirm, that the proposed U-Net based method can detect the outlines of the bricks with high accuracy (above 80%) for any types of walls, significantly surpassing the reference methods, which suffer both from false detection and misdetection effects. While the reference techniques show better results by processing photos of walls with regularly shaped and aligned bricks, their performance is drastically degraded for the irregularly structured stone walls. In summary, we found neither of the two reference methods [8], [9] capable for providing efficient markers for the Watershed process.

Next we evaluate the final output of brick segmentation based on the provided U-Net mask. Fig. 4 shows the results of our algorithm step by step for three sample images. The first row represents the input images, the second one the delineation map by U-Net, the third row displays our brick segmentation output (green lines represent the outlines of the bricks), which is followed by the

Table 1: Evaluation of the delineation step. Comparison of state-of-the-art methods and our proposed U-Net-based approach.

Method	F1-score (%)	Precision (%)	Recall(%)
Riveiro method [9]	23.65	37.04	17.71
Oses method [8]	22.58	39.57	16.91
Proposed method	**81.57**	**81.16**	**82.14**



(a) Raw input images

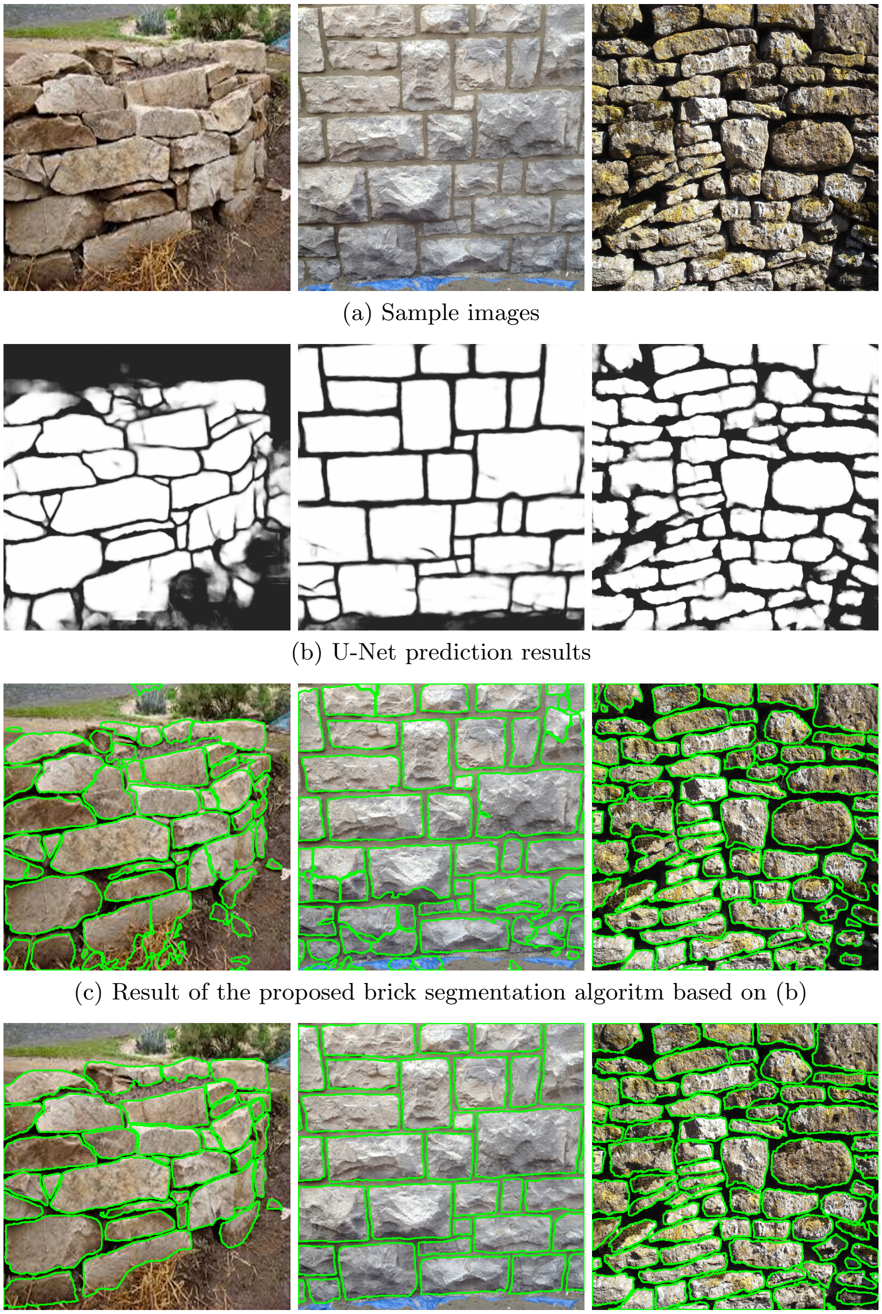
(b) Results of Riveiro's delineation method

(c) Results of Oses' delineation method

(d) Results of the proposed U-Net-based delineation

(e) The ground truth binary brick vs. mortar mask

Fig. 3: Comparison between the-state-of-the-art delineation methods and our method; (a) images; (b) Riveiro's method; (c) Oses' method; (d) Our U-Net based delineation output; (e) Ground truth



(a) Sample images

(b) U-Net prediction results

(c) Result of the proposed brick segmentation algoritm based on (b)

(d) Ground truth

Fig. 4: The results by applying our method step by step for many types of walls; (a) input images; (b) U-Net prediction results; (c) Final segmentation results where the green line is to identify each brick; (d) Ground truth

visualization of the ground truth segmentation. Tabel. 2 shows the quantitative object and pixel level results of the complete workflow. We can conclude that our algorithm provides high quality output for the different wall categories of the test dataset, as the brick and pixel level F1-scores are measured in almost every cases between 76% and 97%. The best results are naturally observed for the ashlar fine subset, which contains high contrasted photos of modern buildings with simple and regular brick layouts. However, as both qualitative and quantitative examples shows, the proposed method can generally handle very different masonry types, and it shows graceful degradation in cases of more challenging samples, such as random masonry.

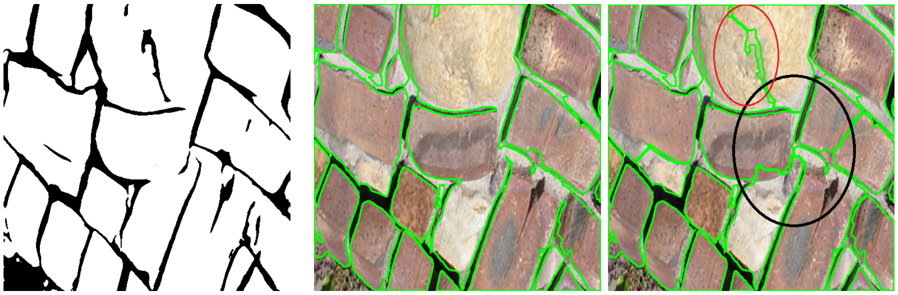
The necessity of applying the marker based Watershed process instead of using a simple connected component analysis (CCA) approach becomes evident by checking Fig. 5 and Table 3. Fig. 5 displays for a sample region the brick segmentation result by CCA and by the proposed Watershed algorithm in parallel. As shown, if some mortar sections are missing or misdetected, neighboring bricks can be erroneously merged into the same object by CCA, while the Watershed approach efficiently handles these situations. Table 3 confirms that such effects may also cause notable differences in quantitative performance parameters, especially for *rough ashlar* walls.

Table 2: Evaluation of brick segmentation. Object (brick) and pixel level precision, recall, F1-score and IOU values for the *augmented* test dataset.

Wall Categories	Number of (augm) images	Recall(%)		Precision(%)		F1-score(%)		IOU(%)
		Brick level	Pixel level	Brick level	Pixel level	Brick level	Pixel level	Pixel level
Random rubble masonry	304	83.80	82.08	77.69	83.03	79.93	81.87	71.31
Square rubble masonry	411	85.23	78.35	69.87	86.10	75.56	78.85	69.91
Dry rubble masonry	375	84.97	85.04	73.54	87.47	77.74	85.36	76.66
Ashlar Fine	268	**97.53**	**96.43**	**97.67**	**92.18**	**97.58**	**94.19**	**89.12**
Ashlar rough	244	81.47	84.19	79.57	78.34	79.87	81.92	69.38
Average	1602	86.38	84.53	78.34	85.67	81.23	83.98	74.88

Table 3: Object (brick) level F1-scores of connected component analysis (CCA) and the proposed Watershed technique for brick segmentation using in both cases our U-Net based delineation maps as input.

The method	Random	Square	Dry	Fine	Rough	Average
CCA	77.46	**77.44**	76.23	95.76	67.02	78.63
Prop. Watershed	**79.93**	75.56	**77.74**	**97.58**	**79.87**	**81.23**



(a) U-Net based mask (b) CCA labeling result (c) Prop. Watershed result

Fig. 5: Comparison of the brick segmentation results with connected component analysis (CCA) and the proposed Watershed technique based on the same U-Net mask.

5 Conclusion

This paper introduced a novel technique for automated brick segmentation in masonry wall images by a joint utilization of the U-Net convolutional neural network and the Watershed segmentation algorithm. The U-Net part provided a high quality delineation map, which enabled efficient marker extraction for the Watershed process. We have shown in a new dataset of diverse masonry photos, that the proposed approach significantly surpasses earlier gradient-driven solutions, and it is largely robust against various noise effects, different illumination conditions, viewpoint and varying masonry types.

Further work will focus on making tests on an extended dataset, and performance comparison of different CNN architectures (like FCN, SegNet, etc.) for the problem. We also plan to expand our studies for 3D point clouds primarily from archaeological sites, with exploiting the advantages of the depth information for brick separation. Another relevant research chapter may deal with wall classification, age or architectural style estimation based on the extracted features.

Acknowledgement

This work was supported by the National Research, Development and Innovation Fund (grants NKFIA K-120233 and KH-125681), and by the Szechenyi 2020 Program (grants EFOP-3.6.2-16-2017-00013 and 3.6.3-VEKOP-16- 2017-00002). Budapest, Hungary.

References

1. Bai, M., Urtasun, R.: Deep watershed transform for instance segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 5221–5229. Honolulu, Hawaii (2017)

2. Bosch, F., Valero, E., Forster, A., Wilson, L., Leslie, A.: Evaluation of historic masonry substrates: towards greater objectivity and efficiency (06 2016). https://doi.org/10.4324/9781315628011-8
3. Fernández, J.L., Lerones, P.M., Casanova, E.Z., García-Bermejo, J.G.: Applying deep learning techniques to cultural heritage images within the inception project. In: EuroMed (2016)
4. Hemmleb, M., Weritz A, F., Schiemenz B, A., Grote C, A., Maierhofer, C.: Multi-spectral data acquisition and processing techniques for damage detection on building surfaces. In: ISPRS Commission V Symposium. pp. 1–6 (1 2006)
5. Kuhn, H.W.: The Hungarian method for the assignment problem. Naval Research Logistic Quarterly **2**, 83–97 (1955)
6. Llamas, J., M. Lerones, P., Medina, R., Zalama, E., Gmez-Garca-Bermejo, J.: Classification of architectural heritage images using deep learning techniques. Applied Sciences **7**, 992 (09 2017). https://doi.org/10.3390/app7100992
7. Muoz, X., Freixenet, J., Cufi, X., Marti, J.: Strategies for image segmentation combining region and boundary information. Pattern Recognition Letters **24**, 375–392 (01 2003). https://doi.org/10.1016/S0167-8655(02)00262-3
8. Oses, N., Dornaika, F., Moujahid, A.: Image-based delineation and classification of built heritage masonry. Remote Sensing **6**(3), 18631889 (Feb 2014). https://doi.org/10.3390/rs6031863, http://dx.doi.org/10.3390/rs6031863
9. Riveiro, B., Conde, B., Gonzalez, H., Arias, P., Caamao, J.: Automatic creation of structural models from point cloud data: The case of masonry structures. ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences **II-3/W5**, 3–9 (8 2015). https://doi.org/10.5194/isprsannals-II-3-W5-3-2015
10. Roerdink, J.B., Meijster, A.: The Watershed transform: Definitions, algorithms and parallelization strategies. Fundam. Inf. **41**(1,2), 187–228 (Apr 2000), http://dl.acm.org/citation.cfm?id=2372488.2372495
11. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention. Lecture Notes in Computer Science, vol. 9351, pp. 234–241. Springer International Publishing (2015). https://doi.org/10.1007/978-3-319-24574-4_28, http://arxiv.org/abs/1505.04597
12. Sithole, G.: Detection of bricks in a masonry wall. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Science **XXXVII**, 567–572 (2008)
13. Tao, Y., Palasek, P., Ling, Z., Patras, I.: Background modelling based on generative unet. In: IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS). pp. 1–6 (Aug 2017). https://doi.org/10.1109/AVSS.2017.8078483
14. Zyuzin, V., Sergey, P., Mukhtarov, A., Chumarnaya, T., Solovyova, O., Bobkova, A., Myasnikov, V.: Identification of the left ventricle endocardial border on two-dimensional ultrasound images using the convolutional neural network unet. In: 2018 Ural Symposium on Biomedical Engineering, Radioelectronics and Information Technology (USBEREIT). pp. 76–78 (May 2018). https://doi.org/10.1109/USBEREIT.2018.8384554