Another Brick in the Wall:
Improving the Assisted Semantic Segmentation of Masonry Walls

G. Pavoni [1] , F. Giuliani [2] , A. De Falco [2] , M. Corsini [1] , F. Ponchio [1] , M. Callieri [1] and P. Cignoni [1]

[1] Visual Computing Lab, ISTI-CNR, Pisa, Italy
[2] Department of Energy, Systems, Territory and Construction Engineering, University of Pisa, Pisa, Italy
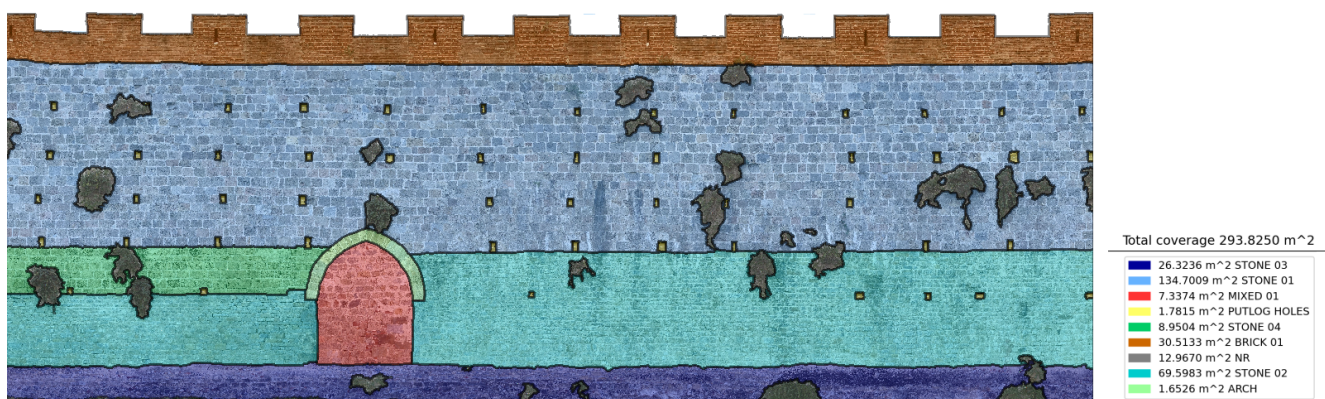

Figure 1: *An ortho-image of an historical architecture (city walls of Pisa,* Vittorio Veneto A *area) segmented in semantic classes representing construction techniques, and their related per-class coverage estimation.*

Abstract

In Architectural Heritage, the masonry's interpretation is an essential instrument for analyzing the construction phases, the assessment of structural properties, and the monitoring of its state of conservation. This work is generally carried out by specialists that, based on visual observation and their knowledge, manually annotate ortho-images of the masonry generated by photogrammetric surveys. This results in vectorial thematic maps segmented according to their construction technique (isolating areas of homogeneous materials/structure/texture) or state of conservation, including degradation areas and damaged parts. This time-consuming manual work, often done with tools that have not been designed for this purpose, represents a bottleneck in the documentation and management workflow and is a severely limiting factor in monitoring large-scale monuments (e.g.city walls). This paper explores the potential of AI-based solutions to improve the efficiency of masonry annotation in Architectural Heritage. This experimentation aims at providing interactive tools that support and empower the current workflow, benefiting from specialists' expertise.

CCS Concepts

• Applied computing → Architecture (buildings); • Computing methodologies → Artificial intelligence; • Human-centered computing → Interactive systems and tools;

1. Context and aims

In the Architectural Heritage (AH) domain, survey-based models and representations of material structures are key tools to address the safety assessment, restoration, and consolidation. The first doc-umentary source for studying historical architectures is the geometry of the building and its construction elements.Traditional geometric surveys and more innovative techniques allow for a complete and extensive metric documentation and knowledge of the

AH at different levels of detail and scale, depending on the scope [BBB*19].

In particular, 2D and 3D photorealistic representations of external masonry surfaces, coming from photogrammetric surveys, allow archaeologists, engineers and conservators to investigate, and document the composition, organisation, construction phases, and damage of walls. The survey of historical wall surfaces is of particular interest in structural engineering to predict the capability of the construction to withstand external actions. This is sometimes preferable to direct experimentation on masonry panels that is difficult to perform, generally expensive, and not always representative of the whole structure. Ancient masonry constructions are often the product of century-old series of transformations that affect the structural homogeneity and the flow of internal forces within the structure. The wall texture bears signs of these changes, as well as past collapses and alterations, and may reveal the quality of the masonry and its attitude to crumble during seismic shaking [BL10]. Furthermore, the strength of the masonry material can be derived using, from the literature, qualitative and quantitative indicators based on the knowledge of materials and block pattern.

The annotation of historical masonry is understood as a process of association between the graphically represented element and any relevant knowledge-based information. As a result of a preliminary diagnostic process, the base representation is covered with a number of patterns, either polygons or regions, and labels that describe the masonry walls (as shown in Figure 1). Two kind of data are usually relevant in the field of AH, namely the characterisation of construction techniques and the identification of the state of conservation [Dog10].

The characterization of the construction techniques consists in the detection of areas with homogeneous material and texture in order to investigate the construction phases and masonry characteristics within the structure [BF10; Dog10]. Significant features are the materials' typology, the geometry of blocks, the filling percentage of joints, and arrangement of units. It is important to remark that a regular arrangement on the external surface may not correspond to a regular section, which is more often extremely irregular. For this reason, the investigation of masonry walls should also account for the thickness and type of section, especially for multi-leaf cases. The same approach is adopted to map degradation, alterations and damage patterns caused by weathering conditions and adverse events [SBJ*12]. Regions with homogeneous phenomena are grouped in classes associated to the presence of vegetation, stains, cracks, rising damp, surface crusts, and spalling of the material.

The conventional annotation approach is based on the manual drawing of the regions over the mappings and, for this reason, it is long and time-consuming. This causes a major bottleneck in the pipeline of creating and updating the documentation, and poses limitation on the ability to manage frequent large-scale monitoring surveys on massive monuments (e.g. city walls). This problem is becoming more and more evident with the availability of off-the-shelf photogrammetric tools that allow the creation of surveys with much lower efforts.

Another issue of these methodologies is the lack of software tools specifically designed for this task. Most of this work is done in image-editing software (like Adobe Photoshop or Illustrator), CAD or GIS tools. 2D CAD tools are probably overkill for this task, with cumbersome interfaces; and while it is true that GIS tools have been specifically created to map information on 2D+ domains, they still are more focused on a different granularity (geographical, and not architectural).

Given the amount of input data involved, the 2D media as a data source, the task of segmenting a domain according to its components and characteristics, it is convenient, nowadays, to consider AI solutions for automatizing the annotation of AH masonry structures (see Section 1.1). Modern Convolutional Neural Network (CNN) could be used to support a specialist in the practical task of tracing the contour of the area he is annotating, and, at the same time, can be employed to automatically segment a whole map, producing a complete annotation of the AH ortho-image. Our aim is not to create an alternative strategy for the AH masonry annotation, but to complement the current workflow with AI-assisted interactive tools and techniques. The idea is to provide tools to support and facilitate the manual annotation, and to automatize some of the large-scale tasks, but always keeping the human experts in-the-loop. In this way, this improved, faster, annotation process is still compatible with what today is the standard workflow in terms of methodologies, input and output data, protocols, and the specialists' expertise.

This paper explores the use of a specific AI-powered tool for the semantic segmentation of orthographic data, TagLab, in the workflow of interpretation and annotation of ortho-images of historical masonry. TagLab is a complete software for the semantic segmentation of 2D orthographic images. It has been developed in the context of analysis of marine biological environments [PEE*19; PCC*20] and, given its generality it can be successfully used for assisted tracing of a generic ortho-image, as it provides high level, content independent, AI-powered tools for the tracing of contours of entities, and a set of specialized editing tools. Additionally, it can be used to train a a semantic segmentation CNN to automatically trace a new ortho-image for specific classification problems.

These features have been used to test the effectiveness of AI methodologies in this task, and to outline a possible integration of these assisting tools in the specialists' consolidated workflow.

1.1. AI-assisted solutions for annotations

In recent years, the performance of semantic segmentation convolutional neural networks has grown enormously. Their progress has led to the parallel development of several platforms dedicated to the data labeling task. Among the many commercial software, we mention Supervisely [https://supervise.ly/], a web-based solution for the data annotation and network training, and LabelBox [https://labelbox.com/].

Generally, labels can be outlined using polygons, bounding boxes, or a positive/negative clicking approach. In a second step, automatic tools refine the rough masks in pixel-wise labels. Castrejón et al. [CKUF17] proposed to speed up polygon tracing using Recursive Neural Network (RNN). However, drawing a polygon always requires 30-40 clicks. Papadopoulos et al. [PUKF17] demonstrate that the task of selecting 4 extreme points (top, bottom, right, and left) is about five times faster than drawing an high-quality bound-

ing box around the object and require a lower cognitive workload. The extremes picking object annotation approach takes an average time of seven seconds.

Starting from the Extreme Clicking approach, Maninis et al. [MCPV18] designed an interactive agnostic segmentation model called Deep Extreme CUT CCN. The Deep Extreme Cut network input a 4-channel data, the RGB image object, and a heatmap which encodes the extremes, and outputs a precise per-pixel label. The heatmap is created by centering four Gaussians on the extreme points indicated by the user. The Deep Extreme Cut can be fine-tuned and learn to outline specific objects.

Another recent click-based solution [FPC*20] builds upon a U-Net [RFB15] architecture and is able to reach an exceptional accuracy (between 95%-99% of mIoU) when a high number of clicks (around 20) is given. The main approach of this network is to not discard previous predictions. Every time a new click is added, the previously obtained masks are given in input together with all the clicks encoded as an image to improve the segmentation.

Some recent works address annotation of the entire image, a task called *full image segmentation*. [AUF18] introduced Fluid Annotation, a labeling methodology that starts by creating a large pool of initial segments. This initial annotation is quickly editable through an *ad-hoc* user interface. The initial segmentation is performed using Mask R-CNN [HGDG17], that is one of the state of the art network for instance segmentation task, carrying out segmentation starting from a bounding box.

This last approaches are not particularly suitable to speed up contour tracing. In this paper, we exploit click-based solutions and other advanced editing instruments to speed up the annotation, as described in details in 2.2.

2. Improving the manual workflow

As described in Section 1, the annotation may happen at multiple levels. In this experimentation we will work on thematic mapping aimed at the characterisation of building techniques, i.e. isolating and annotating those areas with homogeneous material and texture.

We divided the experiment in two parts to evaluate the efficiency of assisted and fully automated tools.

Firstly, following the idea of keeping the experts in-the-loop, by providing tools for assisting their mapping task, we wanted to evaluate how much the use of assisted tracing in TagLab could speed-up the human part of the workflow, with respect to the use of non-specialized tools like Adobe Illustrator or GIS packages. Solving the speed bottleneck is the primary concern of this test, but we are also interested in finding out if the resulting annotations are comparable, in terms of accuracy, with the ones produced with other tools.

As a second step, we tested the use of a segmentation CNN to understand if a completely automatic annotation of this kind of dataset is indeed possible and, if so, what are the performance of the network. This automatic segmentation could really speed-up the annotation process, but probably at the price of some accuracy. For this reason, still inside TagLab, the specialists can use the editing features to correct what has been mis-classified by the CNN.

The two tests are interconnected, as the results of the human assisted annotation are used to train the CNN used in the automatic step.

Ideally, this two-step strategy would perfectly fit in the current annotation workflow, and it would make even more sense when the input dataset is large. The specialists start with the assisted annotation on some representative ortho-images, already gaining an advantage in speed due to the use of a specialized tool. When they have enough data, they can train the CNN on the characteritic of that specific heritage and its specific classes, and then they can automatically annotate the remaining ortho-images with this newly trained CNN. Finally the result of this automatic segmentation may then be corrected using the editing tools.

2.1. Dataset

The photogrammetric survey used in this work covers part of the ancient city walls of Pisa: the north side of the fortification, that was constructed during the XII century, using local materials, techniques and workmanship [BCS11]. The investigated portion is approximately 2 km long, the average height of the walls is 11 m, and the mean thickness is 2.20 m. The structure is made of multi-leaf masonry, with two brick and stone outer-leaves and the inner core of rubble masonry.

Today, the outer side of the town walls is fully accessible and unimpeded, except for localised areas where the sight is hindered by trees. Conversely, the inner side is almost entirely included in private properties and thus inaccessible.

The dimension and extension of the city walls, as well as the particular relation to environment, required the adoption of a rapid photogrammetric surveying workflow, particularly in the data acquisition phase, in order to obtain results of the entire investigated portion in a reasonable time. Photographs were acquired using an iPhone 11 camera having a resolution of 12MP and a 1/2.55-inch sensor, with GPS on. The distance from the wall ranged between 7 m and 12 m depending on the available space and presence of trees, roadways, and fences. Photographs were taken in longitudinal strips with an overlap of 70%, with two bottom-to-top shots wherever the shooting distance was too small to acquire the whole wall height. The acquisition was done over several days at different times, to have the most uniform illumination possible, avoiding too-strong direct light and hard shadows. Data have been processed using Agisoft Metashape to export ortho-images from the generated 3D models. The number of photos in each ortho-image varies between 15 and 30.

The dataset used in this study comprises nine ortho-images (see Table 1) over a total number of 53, with resolution of approximately 3 pixels per cm depending on the acquisition distance.

In spite of the heterogeneous appearance of the walls, seven classes have been initially identified to characterise locally homogeneous areas showing similar construction techniques (Figure 2). The classes consider the lithology, shape and dimension of blocks, the presence of mortar, and finally their arrangement. The latter accounts for the organisation in coursed rows or radial shapes, the even or uneven height of courses, the presence of snecks, and the way units are overlapped. Among these classes, one concerns brick

Sample image				
Class(es) name(s)	**BRICK**	**STONE 01**	**STONE 02**	**STONE 03**
Block	Brick	Dressed stone *Sedimentary Breccia from Asciano*	Roughly dressed stone *Limestone from San Giuliano*	Roughly dressed stones and flints of different types and sizes
Joints	Mortar	Mortar	Mortar	Mortar
Arrangement	Irregular with overlapping units	*Opus pseudo isodomum* almost regular, horizontal courses of nearly even height with overlapping units	*A filaretto* not regular, nearly horizontal courses of variable height with overlapping units	Random, nearly horizontal courses with snecks

Sample image				
Class(es) name(s)	**STONE 04**	**MIXED**	**ARCH**	**NOT RECOGNIZABLE PUTLOG HOLES**
Block	Dressed stone *Yellowish Calcarenite*	Bricks and roughly dressed stones	Dressed wedge-shaped stone (voussoir)	
Joints	Mortar	Mortar	Mortar	
Arrangement	*Opus pseudo isodomum* almost regular, horizontal courses of nearly even height with overlapping units	Random	Radial, units placed with radial joints to create a curved element spanning an opening	Vegetation covering the masonry wall and putlog holes within the stonework

Figure 2: *Semantic classes of the city walls of Pisa. Not recognizable objects and putlog holes (bottom-right) are two separate classes.*

Orthos used in the assisted test and the CNN training	
Contessa Matilde A	15 photos, single vertical shot
Contessa Matilde C	30 photos, single vertical shot
Vittorio Veneto A	16 photos, double vertical shot
Vittorio Veneto B	17 photos, double vertical shot
Vittorio Veneto C	17 photos, double vertical shot
Vittorio Veneto D	17 photos, double vertical shot
Vittorio Veneto E	16 photos, double vertical shot
Orthos used in the automatic test	
Contessa Matilde B	20 photos, single vertical shot
Vittorio Veneto F	21 photos, double vertical shot

Table 1: *Ortho-images used for the assisted and automatic segmentation tests. The images are named according to the road facing the portion of city walls.*

masonry, five describe stone walls, and one regards mixed masonry that is typical of infilled openings and reconstruction works with diverse materials. Two additional classes have been included to map putlog holes and higher plants (i.e. grasses, bushes and even trees) that hinder the recording of masonry.

2.2. Tool description

For the semi-automatic and automatic labeling, we choose Taglab, an Open Source AI-powered annotation tool designed to speed up the annotation and the analysis of large ortho-images. TagLab has been developed by the Visual Computing Lab and is available at https://github.com/cnr-isti-vclab/TagLab.

This all-in-one software covers the entire data labeling and training lifecycle: the dataset preparation, the network training, and the validations of predictions. TagLab integrates different automation degrees (assisted labeling, fully automatic labeling, manual labeling), enabling users non-expert in machine learning to create their annotated datasets and models for automatic image segmentation.

TagLab includes an AI-assisted semi-automatic tracing tool based on the Deep Extreme Cut CNN, called *4-clicks* tool, useful to segment objects with complex contours by simply indicating the

object' extremes. Areas can also be marked out with a manual tracing tool. A set of other image-processing tools are then available to refine, modify and merge/split/carve the segmented areas.

The *Refinement* tool improves the accuracy of jagged boundaries implementing a version of the graph-cut segmentation algorithm [BJ01]. The *Edit Border* tool allows the manual adjustment of boundaries simply by scribbling pixel-level curves intersecting the area being edited. TagLab automatically snaps the beginning and the end of each curve on the old boundaries, filling the inner pixels and removing the outer ones. This tool, based on simple morphological operations on binary masks, is very fast and is an easy way to ensure precise borders.

The *Train-Your-Network* feature, allow the user to create a new classifier by training a DeepLab V3+ network [CZP*18a] with the annotated data. This network has been introduced by Le Chen et al. in 2018 and it is still one of the best performing architectures in terms of accuracy for this task. This CNN follows an "encoder-decoder" structure, using a ResNet-101 as a feature extractor, and natively adopts sparse convolutions to increase neurons' receptive fields. This avoids the input resolution downgrading by usual features pooling operations.

At the end of data annotation, a dedicated interface supports the dataset preparation and the custom fine-tuning of the DeepLab V3+. After the model optimization, users can check the performance's goodness and infer predictions on new images. Taglab also offers a preview of the CNNs predictions on the new data. When the predictions are correct but not completely satisfying, users can exploit a set of image processing tools for quick and accurate editing. In this human-in-the-loop approach, users preserve control over complex operations and, at the same time, being assisted by automatic procedures.

Finally, TagLab automates the extraction of measurements from images (see Figure 1), the exports of tables and histograms, the comparison with multi-temporal inspections, and the use of co-registered DEM information when available. Typically, these workflows require the use of multiple software applications and a computer-science background.

Figure 3: *The 4-clicks segmentation tool in action. (Left) The user marks the extremes points of the area to be segmented. (Middle) The Deep Extreme Cut CNN automatically traces the boundaries. (Right) The Refinement tool can then be used to obtain a more precise segmentation.*

2.3. Assisted annotation pipeline

To evaluate the assisted annotation's effectiveness, we worked with a specialist that already traced other ortho-images of the same city

Figure 4: *The Watershed tool in action. The azure and cyan scribbles mark two areas belonging to specific classes, while the grey scribble marks a "background" area to be ignored. The watershed algorithm transforms the scribbles into segmented areas.*

walls using Adobe Illustrator. After a brief training, the specialist was able to trace the ortho-images independently.

The annotation task exploited by the *4-clicks* tool helps the user in the quick outlining objects such as vegetation, putlog holes, and arches with minimal input (see Figure 3).
This tool works well on "objects", i.e. elements with a clear boundary, but it does not work on large areas, sometimes unbounded, like portions of walls belonging to a class. For this reason, we introduced in TagLab a specific tool for annotating large regions: the *Watershed* tool. The user roughly mark-out areas using scribbles, the tool then applies an adaptation of the watershed segmentation algorithm to segment them (see Figure 4).
Where necessary, the results of both these tracings tools can be locally corrected using the *Refinement* and *Edit Border* tools. The combination of these specialized tools ensured an expedited annotation work.

2.4. Automatized annotation pipeline

The second step of the experimentation exploits the automatic segmentation process. For the model optimization, we use the ortho-images that were manually segmented with the semi-automatic pipeline. The input orthos come from different reconstructions, each one at a slightly different scale. As the pixel size is crucial information to reduce the visual variance and improve the classification performance, all orthos are re-scaled at 1 px = 2.645 mm.

TagLab allows exporting training datasets by slicing large images. During the export, the image and the associated labels are clipped into tiles and saved in separate folders following the partition in three sets: training, validation, and test. In this set-up, we test the CNN performance directly on new ortho-images instead of using a subset of the training data (so, we create only the training and validation folders). Positive performance on new data demonstrates the model's ability to generalize the learned features. TagLab implements different image partition strategies; since classes' distribution is relatively uniform in the longitudinal direction, we choose the left-to-right partition. The seven scaled ortho-images are subdivided into large overlapping tiles of 1026×1026 pixels (scan order: left to right, top to bottom), ending with 1049 labeled tiles; 212 of them are reserved for validation.

The TagLab *Train your network* feature allows the custom fine-tuning of the DeepLab V3+ [CZP*18b] architecture. We perform
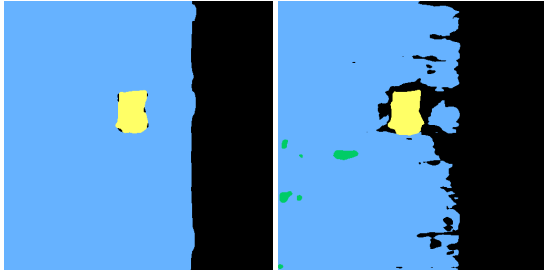
Figure 5: *The model minimizing the Focal Tversky loss outputs (left) cleaner predictions w.r.t the one minimizing the Weighted Cross Entropy loss (right).*

Figure 6: *Segmentation of a caper plant. On left: Adobe Illustrator, on right: TagLab. As explained in Section 2.3, the assisted annotation tools of TagLab allow the tracing of more accurate boundaries in less time.*

geometric augmentation adding small translations and a random scale between $+25\% - 10\%$. After the augmentation, tiles are center-cropped at a resolution of 513×513 pixels, the CNN's input size. The online input normalization subtracts to each tile the dataset per-channel average value. All the pre-trained weights of the DeepLab were let unfrozen, and the learning rate was set lower than the one used during the actual training. Allowing just small updates of weights contrasts the forgetting of high-level features. As an optimizer, we use the Quasi-Hyperbolic Adam optimizer [MY19] with adaptive learning rate decay, an initial learning rate of 10^{-5}, and an L2 penalty of 10^{-4}. We run the model for 110 epochs and a batch size of 32.

Per-class frequencies vary a lot. The BRICKs pixels represent the 7.02% of the total, while STONE 01, the majority class, about the 50%. There are other below-represented classes: the PUTLOG HOLES, with only the 0.51% of pixels, and the Bush (bush and caper bush) with the 3.33%. We trained our model on the BRICK, STONE 01, STONE 02, STONE 03, STONE 04, NR, and PUT-LOG HOLES classes. We discard the MIXED and ARCH classes that are too severely unrepresented in the training dataset.

We mitigate the *class imbalance* following a cost-sensitive approach, acting on the loss function. We compared the performance using a Weighted Cross-Entropy (WCE) loss and a Focal Tversky [AK19] (FT), that auto-balance the classes while training. The model minimizing the FT perform significantly better in term of accuracy and training stability (see Figure 5) . The model fine-tuning required approximately 9*h* using a GPU RTX 2070 with a RAM of 6GB.

3. Results

We tested the assisted annotation pipeline's performance by comparing Illustrator and Taglab on the labeling of the ortho-image *Vittorio Veneto A*. If we do not account for high accuracy, the manual tracing of the boundaries of objects like vegetation and putlog holes takes approximately 15 minutes on Illustrator, while only 9 minutes on TagLab thanks to the *4-clicks* tool. The overall annotation time was 40 minutes with TagLab and about 1 hour and a half with Illustrator. A significant advantage of using TagLab derives from the *Refinement* and *Edit Border* tools that allow boundaries to be more accurate in less time (see Figure 6), whereas Illustrator has less flexible editing options that increase the editing
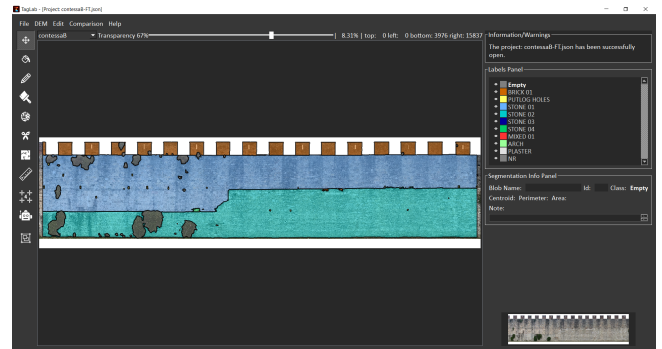
Figure 7: *Automatic masonry predictions on the unseen ortho-image* Contessa Matilde B*, as it appear in the TagLab interface after the automatic classification.*

time. Moreover, Illustrator does not ensure lines to be closed; therefore, further changes are required to create regions and assign a filling pattern associated with the semantic classes.

To evaluate the automatic pipeline, we considered two unlabeled ortho-images (*Vittorio Veneto F* and *Contessa Matilde B*), and we compared the model performance to the two respective human-labeled ground truths. Ground truths were created by annotators running the fully automatic classifier and then editing the predictions through the image processing tools of TagLab. This strategy allows us to measure both the network performance and the time required to correct the predictions. Figure 8 and Figure 7 show the fully automatic prediction of masonry classes exported as a label map. TagLab visualizes labels as polygons superimposed over the ortho-image, Figure 8.

The model reached an accuracy and a mIoU of **0.974** and **0.960** on *Vittorio Veneto F* and of **0.985** and **0.972** on *Contessa Matilde B*. Figure 9 reports the normalized confusion matrix, Figure 10 visualizes the map of human per-pixel editing.

As visible in Figure 10-bottom, in the *Vittorio Veneto F* ortho-image, the STONE 03 class is misclassified with STONE 02 (lower portion). This misclassification error might be due to the low fre-
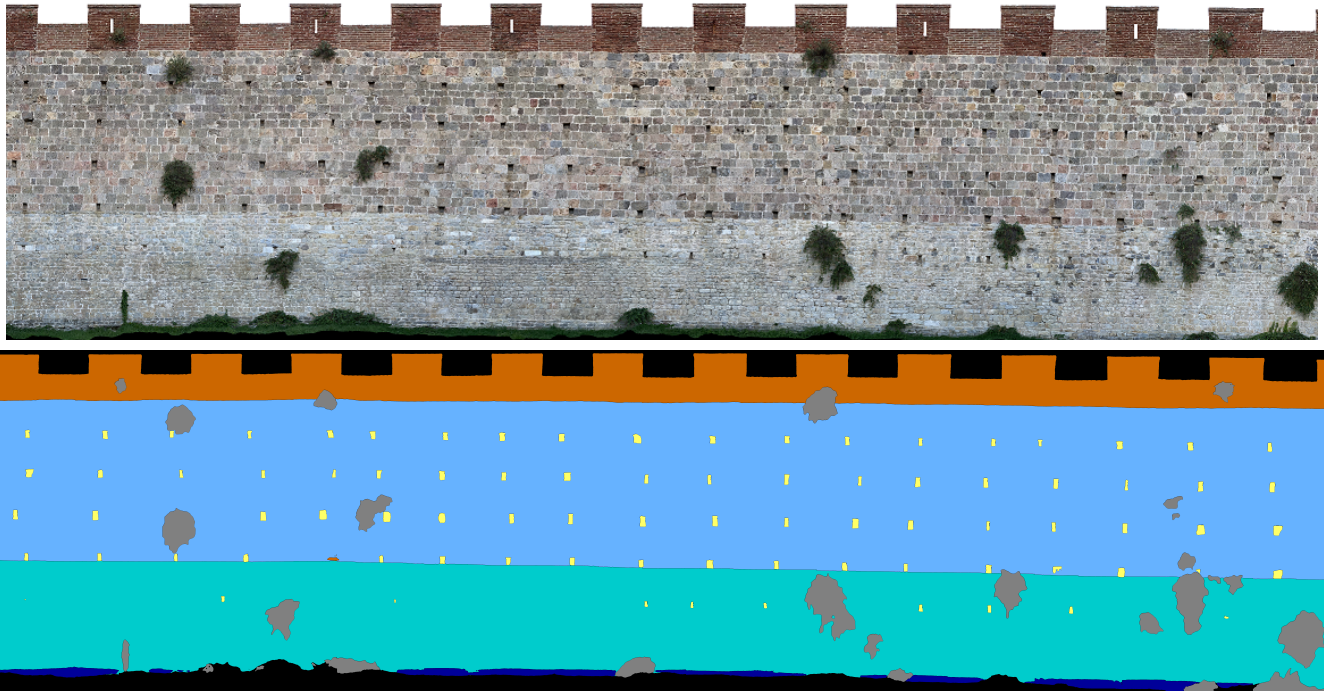
Figure 8: *The* Vittorio Veneto F *ortho-image and the automatic masonry predictions map exported from TagLab as an image.*
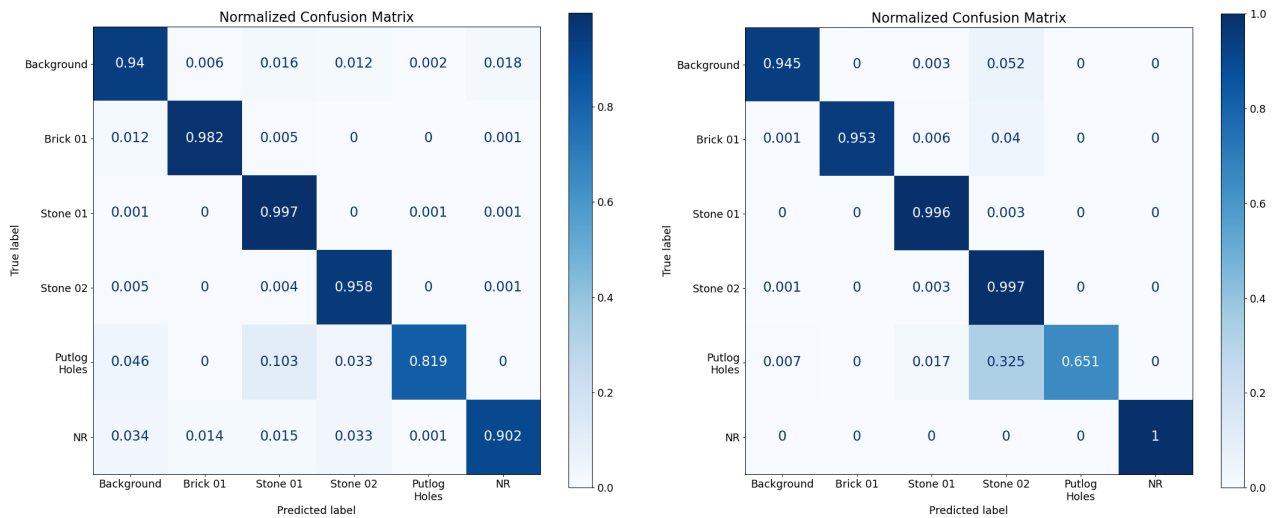

Figure 9: *The confusion matrix of* Vittorio Veneto F *(left) and* Contessa Matilde B *(right). We remark that the ground truth was obtained by editing the automatic predictions. In* Contessa Matilde B *, the annotator considered the annotation of the NR class exact. The STONE 03 and STONE 04 classes are not present in the test ortho-images.*
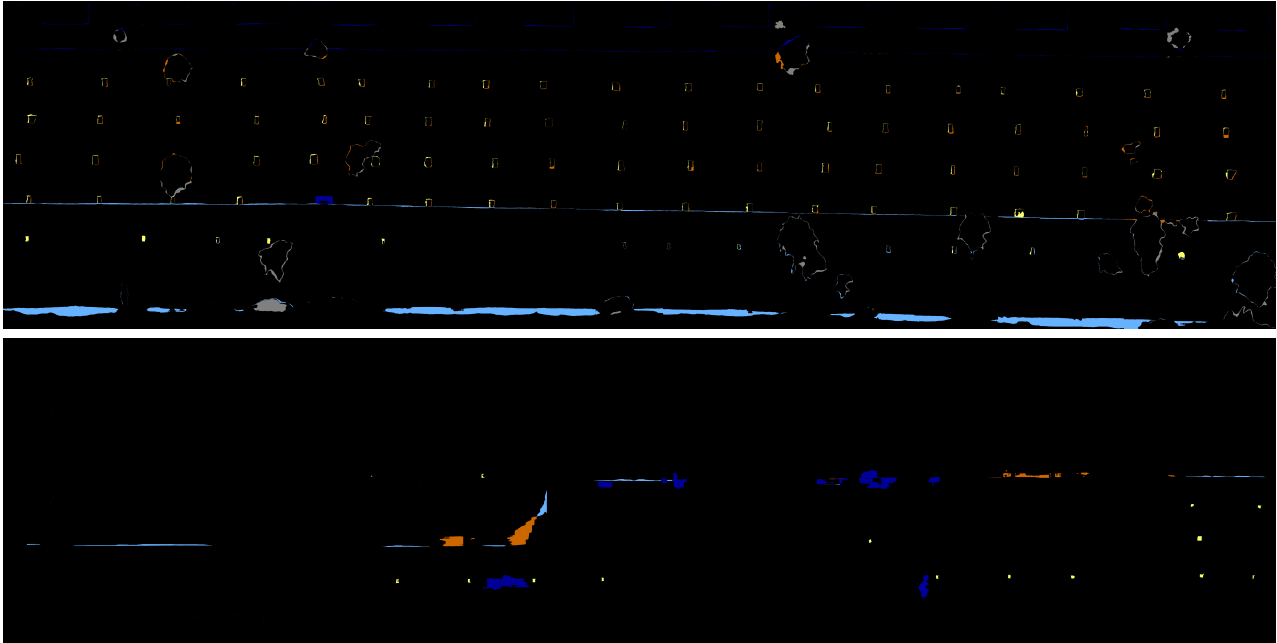
Figure 10: *Pixels edited by users on the automatic labeling of* Vittorio Veneto D *(top) and* Contessa Matilde B *(bottom). This map represents the union of per-class false positives and false negatives.*

quency that the STONE 03 class has in the training dataset. About the other classes, most of the outliers clusters on the boundaries of the objects. The smoother appearance of predicted boundaries is a typical effect of the CNN-based segmentation due to several factors, including the features maps' degradation. Still, the boundaries' accuracy falls below the tolerance of this type of analysis. The *Contessa Matilde B* misclassified areas, visible in Figure 10-top, are of two different types. Blue and Orange areas, detected respectively as belonging to STONE 03 and BRICK classes, actually belong to the MIXED class. However, the MIXED class was not included in the training. Finally, yellow pixels have been mistakenly considered PUTLOG HOLES while missing stones, easy to confuse with PUTLOG HOLES.

The editing of the three automatized annotations took approximately 20 minutes per image. It mainly concerned the redefinition of some of the boundaries between the stone classes and erroneous classes' substitution with the correct one.

4. Conclusions and Future Work

The use of assisted annotation allowed to speed up the manual drawing of boundaries effectively, usually performed using conventional tools.

The results of this experimentation are certainly positive. AI-based tools can be used in this field to support the specialists' work without disrupting their consolidated workflow and providing a relevant speed-up and a satisfactory accuracy of the mapping.

About the assisted solutions for manual annotation, the *4-clicks* tool (which implements Deep Extreme Cut CNN) has proved very

useful in reducing annotation times, as well as the *Refinement* tool. For more radical adjustments, we plan to include in TagLab an AI-based solution for the boundaries editing exploiting positive and negative clicks [FPC*20]. The *Watershed* tool needs to be used carefully to output correct boundaries as it is not sensitive to changes in the image pattern. To accomplish the same task in the future, we plan to introduce a tool inspired by the one-shot texture segmentation [UMBB18], customized to work on masonry annotation.

The automatic segmentation achieved excellent results. The architecture and training methodology were appropriate for optimizing a semantic segmentation model to partitioning masonry according to construction techniques. To improve the results' accuracy, we plan to extend the model to the remaining two classes MIXED and ARCH, adding new positive samples in the training dataset.

Summarizing, when annotating a single map, we can report the following significant time savings: we need one hour and a half using Illustrator, 40 minutes using only the TagLab assisted solutions, and 20 minutes to edit the automatic predictions. Moreover, the use of TagLab improves the accuracy of boundaries and offers the simultaneous estimation of some metric quantities (see Figure 1).

Another common type of analysis in this field is the mapping of degradation and damage patterns, which will automatically perform in the future. This task is certainly more tricky, as phenomena such as cracks, stains, and grime streaking may cross over different underlying materials/texture.

TagLab is a flexible platform that supports multi-modal analysis from different sensors. The current version loads RGB images and co-registered DEMs. Still, its structure also makes it possible to add

additional channels, such as infrared, that could better distinguish structural and extraneous elements such as plants.

Besides partitioning the masonry into semantic classes, a future direction is the automatic extraction of its single constituent elements, like bricks or stones. In [INB19], authors customize a U-Net to detect the initial markers used by the Watershed segmentation algorithm to separate bricks from the mortar. This would allow the extrapolation of useful information such as the minimum and maximum block size, wall leaf connections, horizontally of bed joints, and analyzes masonry's mechanical properties through codified methods like the Masonry Quality Index [BCCD15].

Acknowledgements

References

[AK19] ABRAHAM, NABILA and KHAN, NAIMUL MEFRAZ. "A novel focal tversky loss function with improved attention u-net for lesion segmentation". *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*. IEEE. 2019, 683–687 6.

[AUF18] ANDRILUKA, MYKHAYLO, UIJLINGS, JASPER R. R., and FERRARI, VITTORIO. "Fluid Annotation: A Human-Machine Collaboration Interface for Full Image Annotation". *Proceedings of the 26th ACM International Conference on Multimedia*. MM '18. Seoul, Republic of Korea: ACM, 2018, 1957–1966. ISBN: 978-1-4503-5665-7 3.

[BBB*19] BITELLI, G., BALLETTI, C., BRUMANA, R., et al. "The GAMHer research project for metric documentation of cultural heritage: current developments". *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XLII-2/W11 (2019), 239–246. DOI: 10.5194/isprs-archives-XLII-2-W11-239-2019 2.

[BCCD15] BORRI, ANTONIO, CORRADI, MARCO, CASTORI, GIULIO, and DE MARIA, ALESSANDRO. "A method for the analysis and classification of historic masonry". *Bulletin of Earthquake Engineering* 13.9 (2015), 2647–2665 9.

[BCS11] BEVILACQUA, MARCO GIORGIO, CACIAGLI, COSTANTINO, and SALOTTI, CRISTINA. *Le mura di Pisa: fortificazioni, ammodernamenti e modificazioni dal XII al XIX secolo*. Edizioni ETS, 2011 3.

[BF10] BROGIOLO, GIAN PIETRO and FACCIO, PAOLO. "Stratigrafia e prevenzione". *Archeologia dell'architettura*. Ed. by BROGIOLO, GIAN PIETRO. Vol. XV. All'Insegna del Giglio, 2010, 55–63 2.

[BJ01] BOYKOV, Y. Y. and JOLLY, M. -. "Interactive graph cuts for optimal boundary amp; region segmentation of objects in N-D images". *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*. Vol. 1. July 2001, 105–112 vol.1. DOI: 10.1109/ICCV.2001.937505 5.

[BL10] BOATO, ANNA and LAGOMARSINO, SERGIO. "Stratigrafia e statica". *Archeologia dell'architettura*. Ed. by BROGIOLO, GIAN PIETRO. Vol. XV. All'Insegna del Giglio, 2010, 47–53 2.

[CKUF17] CASTREJÓN, L., KUNDU, K., URTASUN, R., and FIDLER, S. "Annotating Object Instances with a Polygon-RNN". *CVPR*. July 2017, 4485–4493. DOI: 10.1109/CVPR.2017.477 2.

[CZP*18a] CHEN, LIANG-CHIEH, ZHU, YUKUN, PAPANDREOU, GEORGE, et al. "Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation". *CoRR* abs/1802.02611 (2018). arXiv: 1802.02611. URL: http://arxiv.org/abs/1802.02611 5.

[CZP*18b] CHEN, LIANG-CHIEH, ZHU, YUKUN, PAPANDREOU, GEORGE, et al. "Encoder-decoder with atrous separable convolution for semantic image segmentation". *Proceedings of the European conference on computer vision (ECCV)*. 2018, 801–818 5.

[Dog10] DOGLIONI, FRANCESCO. "Leggibilità della costruzione, percorsi di ricerca stratigrafica e restauro". *Archeologia dell'architettura*. Ed. by BROGIOLO, GIAN PIETRO. Vol. XV. All'Insegna del Giglio, 2010, 65–79 2.

[FPC*20] FORTE, MARCO, PRICE, BRIAN, COHEN, SCOTT, et al. "Getting to 99% Accuracy in Interactive Segmentation". *arXiv preprint arXiv:2003.07932* (2020). 3, 8.

[HGDG17] HE, KAIMING, GKIOXARI, GEORGIA, DOLLAR, PIOTR, and GIRSHICK, ROSS. "Mask R-CNN". *The IEEE International Conference on Computer Vision (ICCV)*. Oct. 2017 3.

[INB19] IBRAHIM, YAHYA, NAGY, BALÁZS, and BENEDEK, CSABA. "CNN-Based Watershed Marker Extraction for Brick Segmentation in Masonry Walls". *Image Analysis and Recognition*. Ed. by KARRAY, FAKHRI, CAMPILHO, AURÉLIO, and YU, ALFRED. Cham: Springer International Publishing, 2019, 332–344. ISBN: 978-3-030-27202-9 9.

[MCPV18] MANINIS, K.-K., CAELLES, S., PONT-TUSET, J., and VAN GOOL, L. "Deep Extreme Cut: From Extreme Points to Object Segmentation". *Computer Vision and Pattern Recognition (CVPR)*. 2018 3.

[MY19] MA, JERRY and YARATS, DENIS. "Quasi-hyperbolic momentum and Adam for deep learning". *International Conference on Learning Representations*. 2019 6.

[PCC*20] PAVONI, GAIA, CORSINI, MASSIMILIANO, CALLIERI, MARCO, et al. "On Improving the Training of Models for the Semantic Segmentation of Benthic Communities from Orthographic Imagery". *Remote Sensing* 12.18 (2020), 3106 2.

[PEE*19] PEDERSEN, NICOLE E, EDWARDS, CLINTON B, EYNAUD, YOAN, et al. "The influence of habitat and adults on the spatial distribution of juvenile corals". *Ecography* 42.10 (2019), 1703–1713 2.

[PUKF17] PAPADOPOULOS, D. P., UIJLINGS, J. R. R., KELLER, F., and FERRARI, V. "Extreme Clicking for Efficient Object Annotation". *ICCV 2017*. Oct. 2017, 4940–4949. DOI: 10.1109/ICCV.2017.528 2.

[RFB15] RONNEBERGER, OLAF, FISCHER, PHILIPP, and BROX, THOMAS. "U-Net: Convolutional Networks for Biomedical Image Segmentation". *CoRR* abs/1505.04597 (2015). arXiv: 1505.04597. URL: http://arxiv.org/abs/1505.04597 3.

[SBJ*12] STEFANI, CHIARA, BRUNETAUD, XAVIER, JANVIER-BADOSA, SARAH, et al. "3D Information System for the Digital Documentation and the Monitoring of Stone Alteration". *Progress in Cultural Heritage Preservation*. Ed. by IOANNIDES, MARINOS, FRITSCH, DIETER, LEISSNER, JOHANNA, et al. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, 330–339. ISBN: 978-3-642-34234-9 2.

[UMBB18] USTYUZHANINOV, IVAN, MICHAELIS, CLAUDIO, BRENDEL, WIELAND, and BETHGE, MATTHIAS. "One-shot texture segmentation". *arXiv preprint arXiv:1807.02654* (2018). 8.