On Assisting and Automatizing the Semantic Segmentation of Masonry Walls

GAIA PAVONI, Visual Computing Lab, ISTI-CNR, Italy
FRANCESCA GIULIANI, Department of Civil and Industrial Engineering, University of Pisa, Italy
ANNA DE FALCO, Department of Civil and Industrial Engineering, University of Pisa, Italy
MASSIMILIANO CORSINI, Visual Computing Lab, ISTI-CNR, Italy
FEDERICO PONCHIO, Visual Computing Lab, ISTI-CNR, Italy
MARCO CALLIERI, Visual Computing Lab, ISTI-CNR, Italy
PAOLO CIGNONI, Visual Computing Lab, ISTI-CNR, Italy

In Architectural Heritage, the masonry's interpretation is an essential instrument for analysing 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 vector thematic maps segmented according to their construction technique (isolating areas of homogeneous materials/structure/texture or each individual constituting block of the masonry) 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.*

Additional Key Words and Phrases: Digital Heritage, intelligent systems, interactive semantic segmentation, bricks segmentation, automatic recognition, CNN

ACM Reference Format:
Gaia Pavoni, Francesca Giuliani, Anna De Falco, Massimiliano Corsini, Federico Ponchio, Marco Callieri, and Paolo Cignoni. 2021. On Assisting and Automatizing the Semantic Segmentation of Masonry Walls. *ACM J. Comput. Cult. Herit.* 1, 1, Article 1 (January 2021), 18 pages. https://doi.org/10.1145/3477400

Authors' addresses: Gaia Pavoni, gaia.pavoni@isti.cnr.it, Visual Computing Lab, ISTI-CNR, Via G. Moruzzi 1, Pisa, Italy, 56124; Francesca Giuliani, francesca.giuliani@ing.unipi.it, Department of Civil and Industrial Engineering, University of Pisa, Largo Lucio Lazzarino 1, Pisa, Italy, 56100; Anna De Falco, anna.de.falco@unipi.it, Department of Civil and Industrial Engineering, University of Pisa, Largo Lucio Lazzarino 1, Pisa, Italy, 56100; Massimiliano Corsini, massimiliano.corsini@isti.cnr.it, Visual Computing Lab, ISTI-CNR, Via G. Moruzzi 1, Pisa, Italy, 56124; Federico Ponchio, federico.ponchio@isti.cnr.it, Visual Computing Lab, ISTI-CNR, Via G. Moruzzi 1, Pisa, Italy, 56124; Marco Callieri, marco.callieri@isti.cnr.it, Visual Computing Lab, ISTI-CNR, Via G. Moruzzi 1, Pisa, Italy, 56124; Paolo Cignoni, paolo.cignoni@isti.cnr.it, Visual Computing Lab, ISTI-CNR, Via G. Moruzzi 1, Pisa, Italy, 56124.

Manuscript submitted to ACM

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 documentary source for studying historical architectures is the geometry of the building and its construction elements. Geometrical features form the basis of every operation of conservation aiming at preserving the material and immaterial heritage values, namely the historical building and the traditional body of knowledge and craftsmanship that contributed to its survival over the centuries. 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 [3].

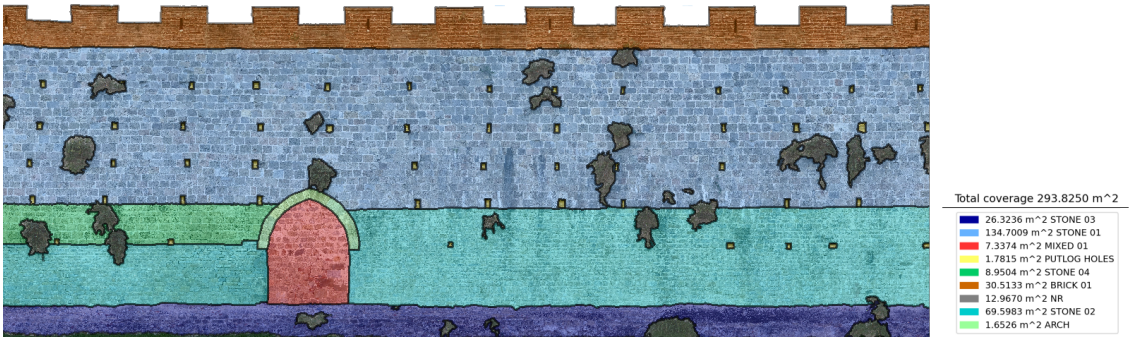

Fig. 1. An ortho-image of an historical architecture (city walls of Pisa, *Vittorio Veneto A* area) segmented in semantic classes representing construction techniques, and their related per-class coverage estimation.

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 [23], 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 [11], and may reveal the quality of the masonry and its attitude to crumble during seismic shaking [4, 6]. 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 Fig. 1). Two kinds of data are usually relevant in the field of AH, namely the characterisation of construction techniques and the identification of the state of conservation [12].

Significant features to identify the construction techniques 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 cross-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 [24]. 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 characterization of the construction techniques may be performed at two different levels of detail. At the level of the whole masonry structure, it consists in the detection of areas with homogeneous material and texture in order to investigate the construction phases and masonry characteristics within the structure [8, 12].
At a finer local level, in areas with a more homogeneous nature, characterization is done by isolating the individual constituent elements: bricks, stones, mortar. This segmentation allows for the extrapolation of useful information such as the block shape and size, horizontality of mortar bed joints, and staggering of vertical mortar joints. Besides, the segmentation of individual blocks is an important source of guidance for physical interventions and consolidation works on historical masonry walls. For instance, once blocks are outlined and enumerated, they can be correctly reassembled once they are dismantled or eventually crumbled, ensuring that the rebuilt form reflects the original. Additionally, tracing the horizontal mortar bed joints may help archaeologists to identify discontinuities, which can be related to different construction ages, or may provide insights on the occurrence of foundation settlements whereas layers are all slanted in the same way.
By segmenting the blocks, it is also possible to gather information about the masonry's mechanical properties through established methods like the Masonry Quality Index (MQI) [5]. This technique consists in the geometrical observation of masonry, the accurate tracing of the constructive elements, and the calculation of a numerical index representative of the masonry quality. The latter is deduced by estimating a set of critical parameters regarding the typological and constructive characteristics that have a direct influence on the structural response of the masonry building under investigation. This computed index can also be used to evaluate the mechanical properties of the masonry, namely the compressive strength, shear strength and Young's modulus, by applying experimentally derived correlation curves [5]. The MQI is particularly interesting in analysing AH assets on which structural testing, especially using invasive or destructive techniques, may not be feasible due to the risk of damaging buildings with high cultural and material value [6, 11].

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 limitations 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).

Nowadays, AI solutions for the 2D semantic segmentation automate the annotation of AH masonry structures (see Section 1.1), facilitating the processing of large amounts of input data. Modern Convolutional Neural Network (CNN) could be used to support a specialist in the practical task of tracing the contour of the area he/she 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 TagLab, a specific AI-powered tool, for the semantic segmentation of orthographic data 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 [20, 21] 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 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 convolutional neural networks in the semantic segmentation task has grown enormously. Their progress has led to the parallel development of several platforms dedicated to the data labelling task. Among the many commercial software, we mention Supervisely, a web-based solution for the data annotation and network training, and LabelBox. Generally, labels can be outlined using polygons, bounding boxes, or a points-clicking approach. Castrejón et al. [9] proposed to speed up polygon tracing using Recursive Neural Network (RNN). However, drawing a polygon always requires multiple clicks, while bounding box annotations are undoubtedly faster. Starting from an initial bounding box, a precise object segmentation can be carried out through several methods; most of them use different declinations of the Mask R-CNN [14].

Papadopoulos et al. [19] demonstrate that the task of selecting 4 extreme points (top, bottom, right, and left) is about five times faster than drawing an high-quality bounding box around the object and require a lower cognitive workload. The annotation of an object by picking its extremities takes an average time of around seven seconds. Starting from the Extreme Clicking approach, Maninis et al. [18] designed an interactive agnostic segmentation model called Deep Extreme Cut CCN. The Deep Extreme Cut network uses as input a 4-channel data, the RGB object image and a heatmap which encodes its extremities, and outputs a precise per-pixel label.

A recent click-based solution [13], builds upon a U-Net [22] architecture, scores an exceptional accuracy (between 95%-99% of mIoU) when a high number of clicks (around 20) is given. This model works iteratively; every time a new click is added, previously masks are given in input, and all the clicks are encoded as an image to improve the

segmentation. Always exploiting the segmentation masks from previous steps, but employing a lower number of positive/negative clicks, [16] reaches a remarkable level of accuracy.

Concerning the walls' segmentation into individual blocks, recent works such as [15] uses the U-Net [22] architecture. This promising strategy implies that the CNN, known for its applications in medical imaging, is optimized to work on masonry data.

In this paper, we exploit two different click-based solutions and other advanced editing instruments to speed up the annotation of objects, as detailed in Section 2.2. Additionally, since there are no publicly available datasets to train a neural network for stones/bricks segmentation, we implemented a *bricks segmentation* tool. This tool allows reaching two goals: to segment individual bricks/stones very quickly w.r.t manual segmentation and to create a training dataset for the future development of a specific CNN for this task.

2 IMPROVING THE MANUAL WORKFLOW

As described in Section 1, the annotation may happen at multiple levels. In this experimentation we will firstly work at a higher level, on a mapping aimed at the characterisation of building techniques: i.e. isolating and annotating those areas with homogeneous material and texture. Then, we will work inside those mapped areas, to identify and segment the individual stones/bricks.

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 next 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 the performance of the network is. 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 stages of assisted and automatic 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 characteristic 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 [2]. 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.

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.

In spite of the heterogeneous appearance of the walls, seven classes have been initially identified to characterise locally homogeneous areas showing similar construction techniques (Fig. 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 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 plants (i.e. grasses, bushes and even trees) that hinder the recording of masonry.

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

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

2.2 Tool description

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

TagLab implements two AI-assisted interactive annotation tools: the *4-clicks* tool, based on the Deep Extreme Cut CNN [18], and the *positive/negative clicks* tool, based on the CNN recently presented in [16]. Both models have been fine-tuned to works on jagged-shaped objects, exploiting a manually labelled, highly accurate dataset. Using the *4-clicks* tool, the user traces the objects' boundaries by indicating the four extreme points (Fig. 3). With the *positive/negative clicks* tool, the user outlines an object's by picking an arbitrary number of inner (positive) and outer (negative) points. In most cases, a single internal point is sufficient to segment an area (Fig. 4). This last tool works both as a tracing tool, creating

a new segmented entity, and as an editing/refining tool, adding and removing pieces from an existing segmented entity (Fig. 4). These two intelligent resources can speed up the segmentation of a large number of objects. However, for extremely complex cases, TagLab also offers a manual per-pixel tracing tool that gives the user full control over the outlines.

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

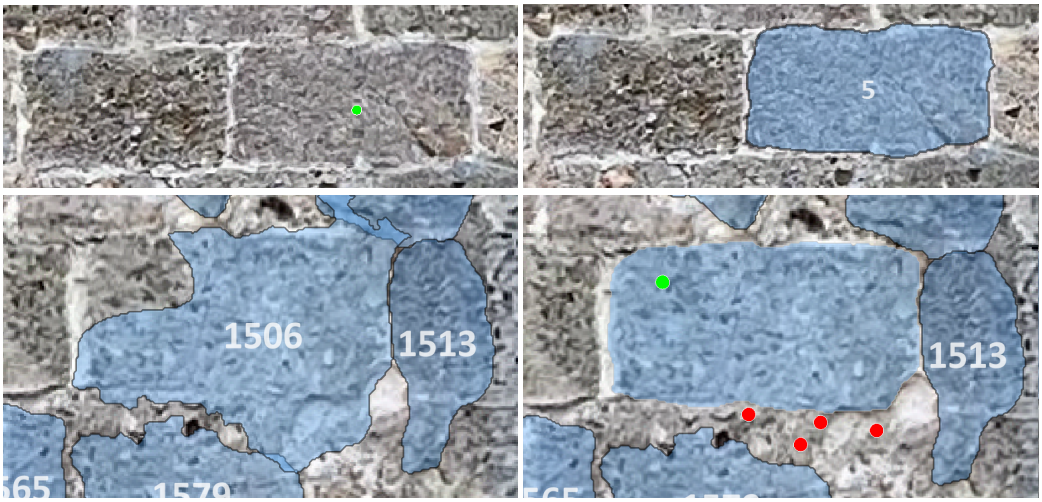

Fig. 4. The *positive/negative clicks* tool in action. A single internal click is often enough to segment a brick (Top). The tool also allows the user to correct an existing segmentation placing negative clicks (to exclude regions) in red and positive clicks (to include regions) in green (Bottom).

A set of 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 [7]. 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. Based on simple morphological operations on binary masks, this custom tool ensures the precise editing of borders in short times.

After the user has annotated a dataset specific for its needs, the *Train-Your-Network* feature may be used to train a new classifier specialized for this specific annotated data, with TagLab taking care of all the steps of data preparation. To generate this specialized classifier, TagLab uses a DeepLab V3+ architecture [10]. Introduced by Le Chen et al. in 2018, it is still one of the best performing semantic segmentation network in terms of accuracy. 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 features pooling operations.
After the training, the users can evaluate the performance of the specialized classifier using numerical and graphical feedback, and decide if the model needs more tweaking or it is ready to be used.
The specialized classifier can be used in TagLab to infer predictions on new images (on a single ortho, or in batch on large datasets). These predictions may then be manually corrected by the user with the same AI-assisted and image processing tools described above, reaching a segmentation accuracy comparable with human experts.
In this human-in-the-loop approach, the user retains full control over each step, fully exploiting his knowledge of the field, but at the same time is assisted by automatic procedures that speed-up its annotation/correction work or that automatize cumbersome tasks.

Finally, TagLab automates the extraction of measurements from annotated images (see Fig. 1), the exporting 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.

2.3 Semantic Classes - Assisted annotation

To evaluate the assisted annotation's effectiveness, we worked with a specialist that already traced other ortho-images of the same city walls using Adobe Illustrator. After a brief training, the specialist was able to trace the ortho-images independently.

The annotation task exploited the *4-clicks* tool, that helps the user in the quick outlining objects such as vegetation, putlog holes, and arches with minimal input (see Fig. 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 Fig. 5).
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 Semantic Classes - Automatized annotation

After the experimentation with the assisted tools, the next step was focused on the automatic segmentation process. For the model optimization, we use the ortho-images that were manually segmented with the semi-automatic pipeline. The

Fig. 5. 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.

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 sets). 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 a 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 labelled tiles; 212 of them are reserved for validation.

We perform the 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 [17] 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 BRICK class 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 PUTLOG 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 [1] (FT), that auto-balance the classes while training. The model minimizing the FT perform significantly better in term of accuracy and training stability. The model fine-tuning required approximately $9h$ using a GPU RTX 2070 with a RAM of 6GB.

2.5 Individual Blocks - Assisted annotation

As detailed in Section 1, after obtaining a segmentation of the masonry structure in its semantic classes, it is possible to work at a finer level, and annotate each individual stone/brick inside each semantic class macro-area. Also in this part of the experimentation, our aim was to test the effectiveness of the assisted tools available in TagLab, and how to effectively fit them in the existing manual pipeline. The change in granularity, cardinality and size of the areas to annotate required a change in the tools used. While tracing *few* elements using the *4-clicks* and *positive/negative clicks* tools is definitely possible and effective, this strategy does not scale up well for larger areas, where there might be thousands of individual building blocks to trace. For this reason, we introduced in TagLab a specialized element-tracing tool, the *bricks segmentation* tool (see Fig. 6) that quickly segments individual bricks/stones over a large area. The individual constructive elements may exhibit very different shapes, sizes, and visual aspects; some blocks can be small with irregular shapes, even rectangular bricks may vary a lot in sizes in the same area, and so on. To take into account these differences, the *bricks segmentation* tool provides two dedicated algorithms: one more effective for blocks with a more rectangular shape, and another one for blocks with a more irregular one. The user should choose an approximate value for the minimum and maximum distance between the individual elements (easy measurable through the ruler tool) and choose between the two different image processing algorithms. A single thresholding slider helps the user to adjust the algorithm to the different constructive elements. Fig. 6 shows an ideal thresholding output; each brick/stone is marked with one/two red points. The position of these points is estimated using the edges extracted by one of the dedicated algorithms (also the extracted edges are shown in the output preview). TagLab uses those seed points as positive inputs for the *positive-negative clicks* CNN and generates the negative click taking into account the given distances between elements, outputting the accurate segmentation of the wall into single constituents (see Section 3.2). The resulting segmentation of individual blocks can eventually be refined using interactive editing tools. The whole assisted process, when compared to the manual annotation, is considerably faster.

3 RESULTS

3.1 Semantic Classes

We tested the assisted annotation pipeline's performance by comparing Illustrator and Taglab on the labelling 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 Fig. 7), whereas Illustrator has less flexible editing options that increase the editing time. Additionally, 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 unlabelled ortho-images (*Vittorio Veneto F* and *Contessa Matilde B*), and we compared the model performance to the two respective human-labelled 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. Fig. 8 and Fig. 9 show the fully automatic prediction of masonry classes exported as a label map. TagLab visualizes labels as polygons superimposed over the ortho-image (Fig. 9).

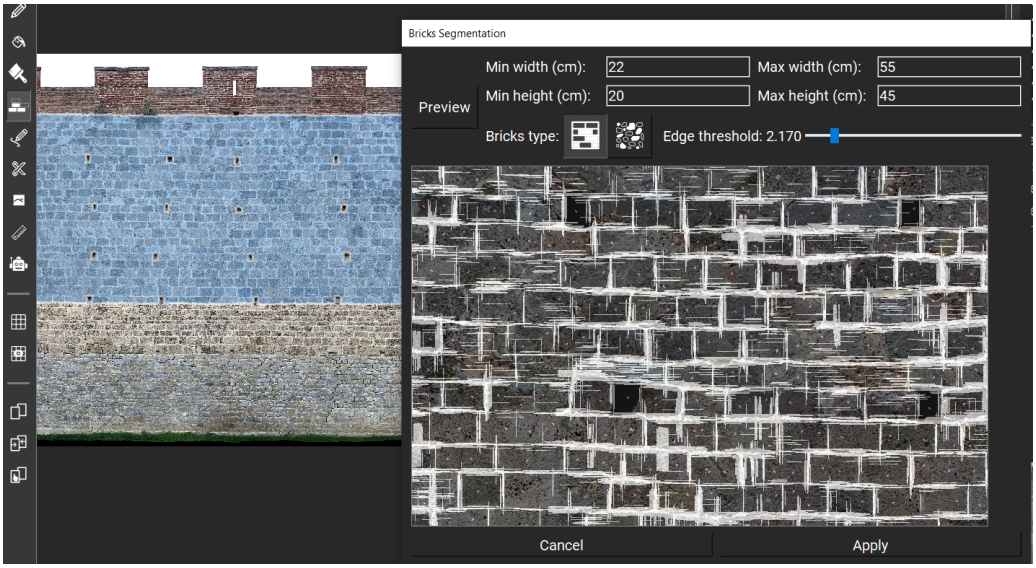

Fig. 6. The *bricks segmentation* tool in action. This tool is applied to single semantic areas; in this image to the STONE 01 class. Once the measurements have been provided, the user selects the appropriate algorithm, in this case the one for more regular bricks, and adjusts the threshold until obtaining one/two red points per brick. The results of this parameters are visible in Fig. 12

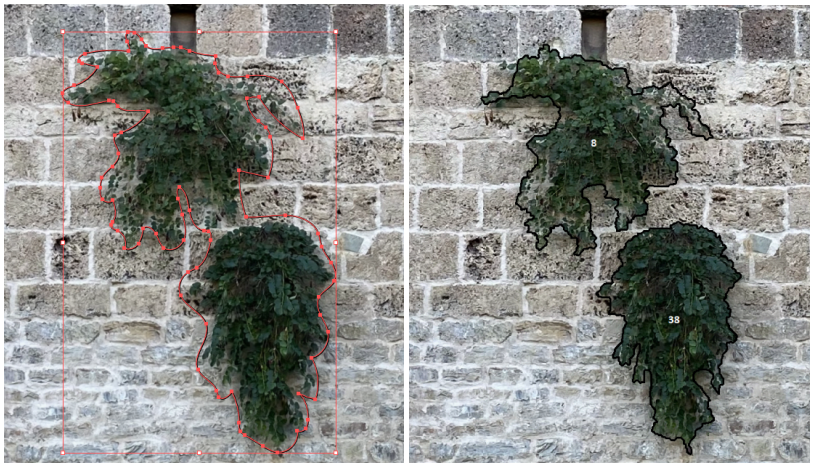

Fig. 7. Segmentation of a caper plant. On left: Adobe Illustrator, on right: TagLab. As explained in Section 2.3, the assisted annotation tools of TagLab allow the tracing of more accurate boundaries in less time.

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*. Fig. 10 reports the normalized confusion matrix, Fig. 11 visualizes the map of human per-pixel editing.

As visible in Fig. 11-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 frequency that the STONE 03 class has in the

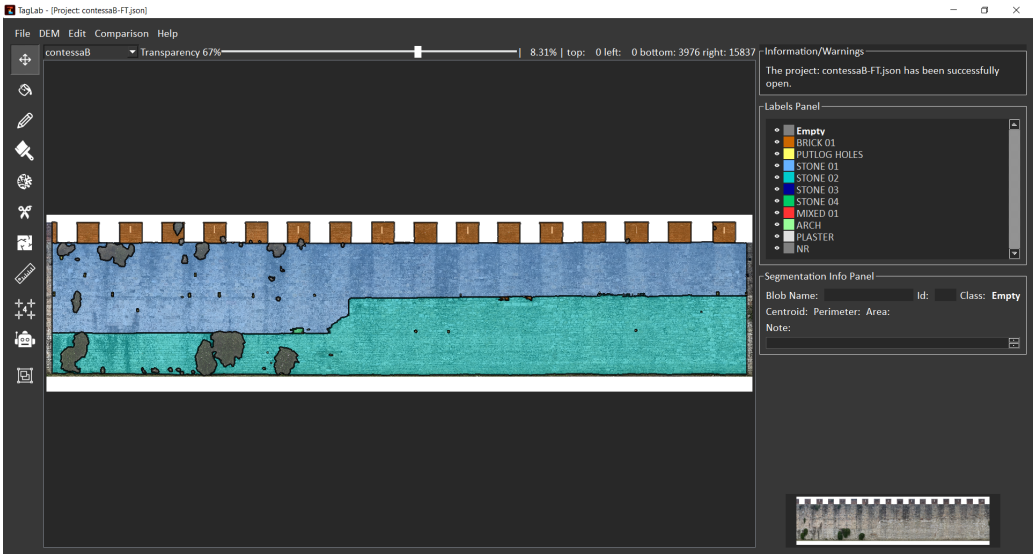

Fig. 8. Automatic masonry predictions on a new unlabelled ortho-image *Contessa Matilde B*, as it appear in the TagLab interface after the automatic classification.

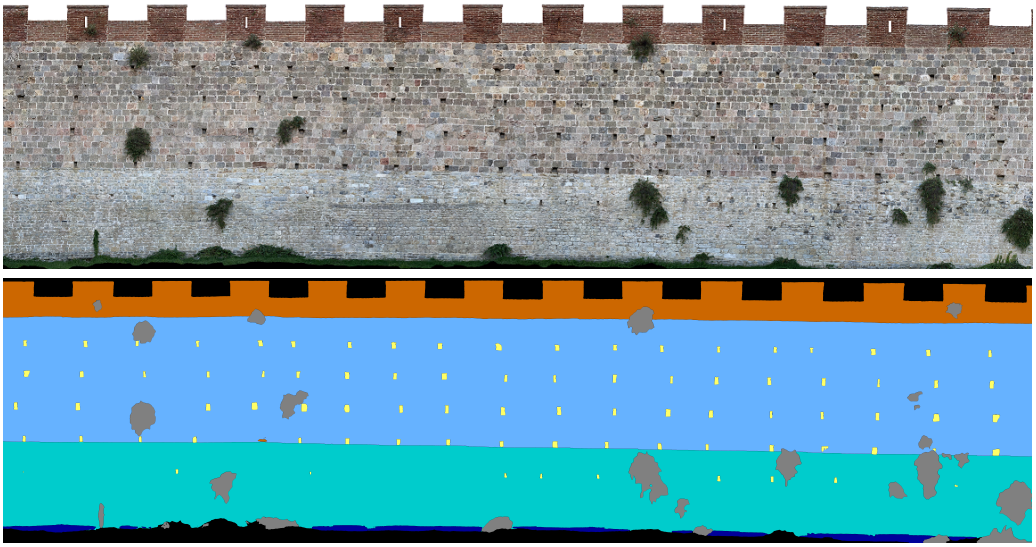

Fig. 9. The *Vittorio Veneto F* ortho-image and the automatic masonry predictions map exported from TagLab as an image.

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 Fig. 11-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

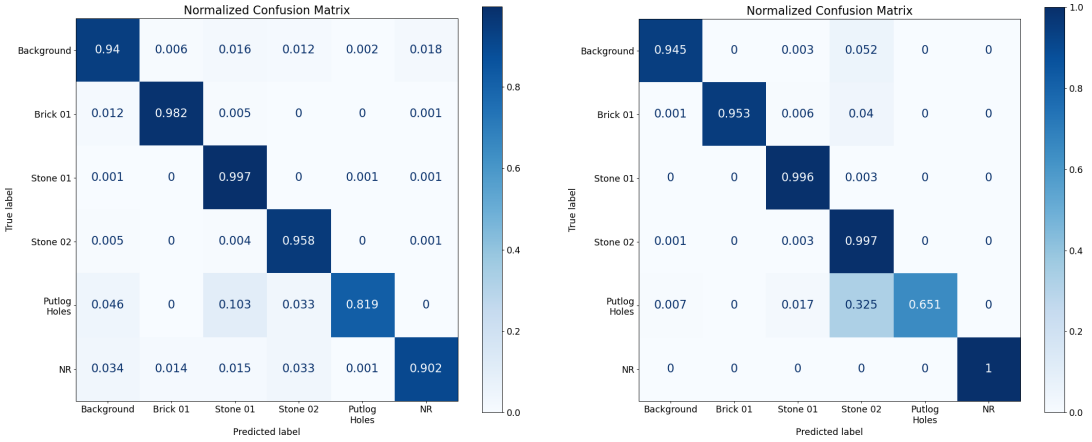

Fig. 10. The confusion matrix of *Vittorio Veneto F* (left) and *Contessa Matilde B* (right). We remark that the ground truth was obtained by editing the automatic predictions; so in *Contessa Matilde B*, the annotator evaluated the annotation of the NR class totally correct. STONE 03 and STONE 04 classes were not present in the test ortho-images.

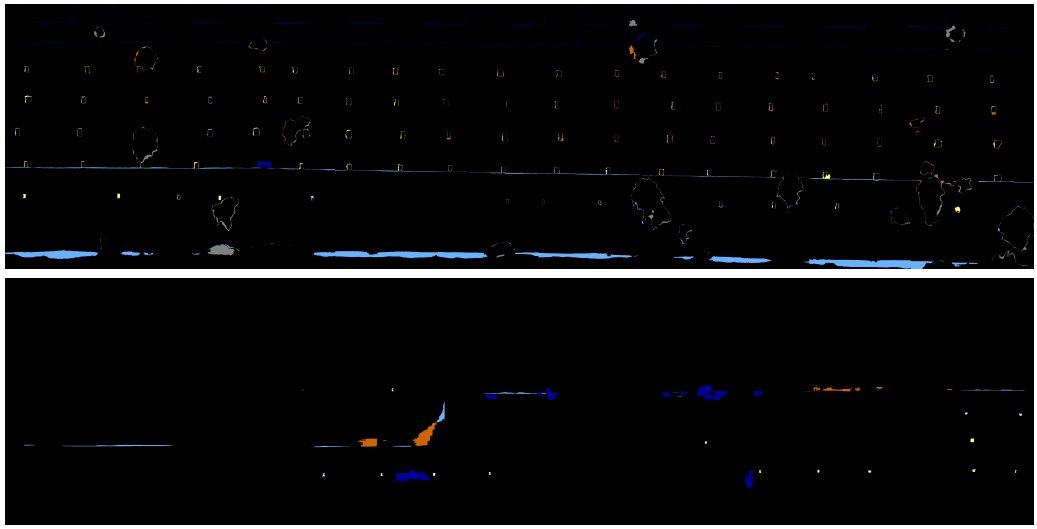

Fig. 11. Pixels edited by users on the automatic labelling of *Vittorio Veneto D* (top) and *Contessa Matilde B* (bottom). This map represents the union of per-class false positives and false negatives.

class was not included in the training. Finally, yellow pixels have been mistakenly considered PUTLOG HOLES while they actually were missing stones.

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.

3.2 Individual Blocks

The availability of multiple tracing and editing tools played an important role in improving this tedious manual task. The *bricks segmentation* tool succeed in the majority of the semantic areas, while the few missing or inaccurate elements can be rapidly edited using the *positive/negative clicks* tool. To facilitate the ortho inspection, TagLab allows for activating a grid (of assignable size and position) with cells that can be marked and annotated, to keep track of the areas already visited or edited.
In terms of speed and scalability, the *bricks segmentation* tool proved to be effective. In Fig. 12, of the *Vittorio Veneto D* map, the tool segment 1839 stones, over a surface of around 176 square meters. Once the right thresholding value has been provided, the automatic tracing took around 10 minutes, a vast improvement over a completely manual mapping.

The resulting segmentation could directly be used for an autoptic examination and measurement of the masonry, inside Taglab. The visualization of the segmented elements over the orthoimage helps the identification of discontinuities in the arrangement of units and those areas where the geometry of blocks is locally different with respect to the rest of the semantic class. The detection of these variations may often be hard, especially when conducting rapid surveying or when the close scrutiny of the whole surface is challenging due to logistic issues (e.g., on the top of city walls). Nonetheless, their identification and analysis are usually significant for fully understanding the construction process and hypothesizing the presence of the alterations, thus orienting archival evidence.

Along with providing rapid and accurate results, TagLab ensures a greater usability by enabling the extrapolation of significant numerical data from segmented masonry walls. Specialists often rely on a series of values calculated over the segmented elements, such as statistics for the height, width, and aspect ratio of the stones; tracings of the individual rows of elements; detected discontinuities of the layout. As the segmented bricks/stones are stored in TagLab as vector data, all these quantities can be easily generated: an example are the statistics of the blocks' size shown in Fig. 13. For each semantic class, the histogram correlates the area of elements identified for each class (in cm^2) on the abscissa with the number of elements on the ordinate. We can easily state that the dimensions of blocks in the four masonry classes are quite diverse, with the elements tagged as BRICK and STONE 02 having a smaller area with respect to the others. Their distribution is left-skewed, suggesting a greater uniformity in the dimensions. For instance, bricks are characterised by peak values ranging between 90-150 cm^2, which are realistic since the wall portion is formed laying headers and stretchers and bricks were usually produced in standard shapes even if handmade.
The dimension of blocks in the classes STONE 01 and STONE 04 present more disperse values with a roughly symmetric distribution. STONE 04 covers a smaller portion of the wall (and the ortho) under investigation, hence the number of units is reasonably lower than other classes. The histogram makes it possible the identification of outliers and long-tailed data to be compared with the real appearance of the wall. Numerical data may eventually be used as corroborative evidence for the autoptic examination if the blocks associated with these values tend to be spatially localized in a limited portion of the ortho.

4 CONCLUSIONS AND FUTURE WORK

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.

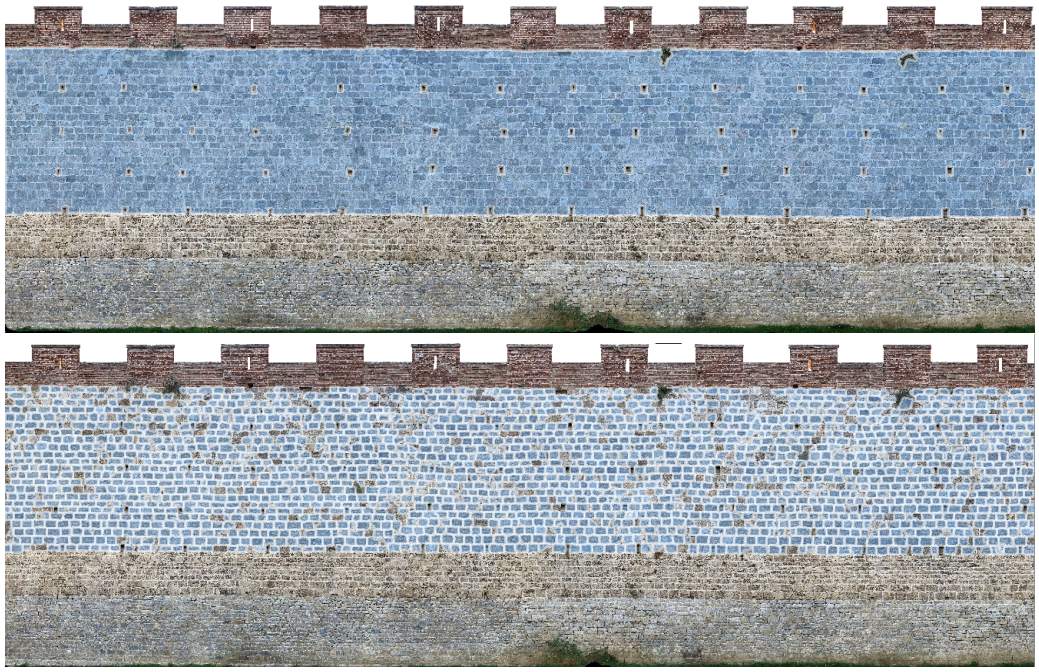

Fig. 12. The semantic class STONE 01 covers 176 square meters in the *Vittorio Veneto D* map (Top). The *bricks segmentation* tool traced 1839 stones for this area (Bottom). Most of the stones are correctly marked; the missing or partially-traced bricks can later be fixed by using the *positive/negative clicks* tool.

Fig. 13. The *bricks segmentation* tool automatically traced 6169 bricks in the the ortho *Vittorio Veneto A*, working on the STONE 01, STONE 02, STONE 04 and BRICK semantic classes, previously detected in Fig. 1. STONE 02 has hardly detectable bricks due to their small size and the white mortar's presence that confounds the outlines. The histogram shows the area of elements identified for each class (in cm^2) on the abscissa while the ordinate reports the number of elements.

The assisted annotation approach was able to speed up considerably the manual drawing of boundaries, usually performed with conventional software tools. The AI-powered *4-clicks* and *positive/negative clicks* tools for tracing areas, as well as the *Refinement* and *Edit Border* tools for modifying them, proved to be extremely effective in reducing the annotation times, as shown in the tests. 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 [25], customized to work on masonry annotation. The specialized *bricks*

segmentation tool worked fairly well, allowing to segment the blocks of a large areas of the masonry in a single step. However, some manual tweaking of parameters remains necessary. Given the vast speedup, we didn't deem necessary the execution of extensive tests. Consider for example that the annotation of the STONE 01 class (Fig. 12) involves the tracing of almost two thousands entities: something that would have required around 10-15 hours (20-30 secs per element with no interruption) instead of just 10 minutes with the proposed system.
By the joint use of the *bricks segmentation* tool and the interactive user corrections (through the *positive/negative clicks* tool), we are able to produce a clean bricks segmentation dataset in a reasonable time.
In the future, we plan to optimize an instance segmentation network to accomplish the segmentation of individual stones/bricks in a fully automatic way.

The automatic segmentation of semantic classes 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. To summarize, when annotating semantic macro-areas in 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 editing the automatic predictions. Additionally, the use of TagLab improves the accuracy of boundaries and offers the simultaneous estimation of some metric quantities (see Fig. 1).

Another common type of analysis in this field is the mapping of degradation and damage patterns, which we will automatically perform in the future. This task is certainly trickier, 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.

ACKNOWLEDGMENTS

This work has been partially supported by the Innovation for Data Elaboration in Heritage Areas - IDEHA project (code number ARS01_00421), National Research Program, MIUR.

REFERENCES

[1] Nabila Abraham and Naimul Mefraz Khan. 2019. A novel focal tversky loss function with improved attention u-net for lesion segmentation. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*. IEEE, 683–687.

[2] Marco Giorgio Bevilacqua, Costantino Caciagli, and Cristina Salotti. 2011. *Le mura di Pisa: fortificazioni, ammodernamenti e modificazioni dal XII al XIX secolo.* Edizioni ETS.

[3] G. Bitelli, C. Balletti, R. Brumana, L. Barazzetti, M. G. D'Urso, F. Rinaudo, and G. Tucci. 2019. 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. https://doi.org/10.5194/isprs-archives-XLII-2-W11-239-2019

[4] Anna Boato and Sergio Lagomarsino. 2010. Stratigrafia e statica. In *Archeologia dell'architettura*, Gian Pietro Brogiolo (Ed.), Vol. XV. All'Insegna del Giglio, 47–53.

[5] Antonio Borri, Marco Corradi, Giulio Castori, and Alessandro De Maria. 2015. A method for the analysis and classification of historic masonry. *Bulletin of Earthquake Engineering* 13, 9 (2015), 2647–2665.

[6] Antonio Borri, Marco Corradi, and Alessandro De Maria. 2020. The Failure of Masonry Walls by Disaggregation and the Masonry Quality Index. *Heritage* 3, 4 (2020), 1162–1198.

[7] Y. Y. Boykov and M. . Jolly. 2001. Interactive graph cuts for optimal boundary amp; region segmentation of objects in N-D images. In *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, Vol. 1. 105–112 vol.1. https://doi.org/10.1109/ICCV.2001.937505

[8] Gian Pietro Brogiolo and Paolo Faccio. 2010. Stratigrafia e prevenzione. In *Archeologia dell'architettura*, Gian Pietro Brogiolo (Ed.), Vol. XV. All'Insegna del Giglio, 55–63.

[9] L. Castrejón, K. Kundu, R. Urtasun, and S. Fidler. 2017. Annotating Object Instances with a Polygon-RNN. In *CVPR*. 4485–4493. https://doi.org/10.1109/CVPR.2017.477

[10] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. 2018. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. *CoRR* abs/1802.02611 (2018). arXiv:1802.02611 http://arxiv.org/abs/1802.02611

[11] Ludovico Dipasquale, Luisa Rovero, and Fabio Fratini. 2020. Ancient stone masonry constructions. In *Nonconventional and Vernacular Construction Materials*. Elsevier, 403–435.

[12] Francesco Doglioni. 2010. Leggibilità della costruzione, percorsi di ricerca stratigrafica e restauro. In *Archeologia dell'architettura*, Gian Pietro Brogiolo (Ed.), Vol. XV. All'Insegna del Giglio, 65–79.

[13] Marco Forte, Brian Price, Scott Cohen, Ning Xu, and François Pitié. 2020. Getting to 99% Accuracy in Interactive Segmentation. *arXiv preprint arXiv:2003.07932* (2020).

[14] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross B. Girshick. 2017. Mask R-CNN. *CoRR* abs/1703.06870 (2017). arXiv:1703.06870 http://arxiv.org/abs/1703.06870

[15] Yahya Ibrahim, Balázs Nagy, and Csaba Benedek. 2019. CNN-Based Watershed Marker Extraction for Brick Segmentation in Masonry Walls. In *Image Analysis and Recognition*, Fakhri Karray, Aurélio Campilho, and Alfred Yu (Eds.). Springer International Publishing, Cham, 332–344.

[16] Anton Konushin Konstantin Sofiiuk, Ilia Petrov. 2021. Reviving Iterative Training with Mask Guidance for Interactive Segmentation. *arXiv preprint arXiv:2102.06583* (2021).

[17] Jerry Ma and Denis Yarats. 2019. Quasi-hyperbolic momentum and Adam for deep learning. In *International Conference on Learning Representations*.

[18] K.-K. Maninis, S. Caelles, J. Pont-Tuset, and L. Van Gool. 2018. Deep Extreme Cut: From Extreme Points to Object Segmentation. In *Computer Vision and Pattern Recognition (CVPR)*.

[19] D. P. Papadopoulos, J. R. R. Uijlings, F. Keller, and V. Ferrari. 2017. Extreme Clicking for Efficient Object Annotation. In *ICCV 2017*. 4940–4949. https://doi.org/10.1109/ICCV.2017.528

[20] Gaia Pavoni, Massimiliano Corsini, Marco Callieri, Giuseppe Fiameni, Clinton Edwards, and Paolo Cignoni. 2020. On Improving the Training of Models for the Semantic Segmentation of Benthic Communities from Orthographic Imagery. *Remote Sensing* 12, 18 (2020), 3106.

[21] Nicole E Pedersen, Clinton B Edwards, Yoan Eynaud, Arthur CR Gleason, Jennifer E Smith, and Stuart A Sandin. 2019. The influence of habitat and adults on the spatial distribution of juvenile corals. *Ecography* 42, 10 (2019), 1703–1713.

[22] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. *CoRR* abs/1505.04597 (2015). arXiv:1505.04597 http://arxiv.org/abs/1505.04597

[23] M.P. Schuller, R.H. Atkinson, and J.L. Noland. 1995. Structural evaluation of historic masonry buildings. *APT Bulletin: The Journal of Preservation Technology* 26, 2/3 (1995), 51–61.

[24] Chiara Stefani, Xavier Brunetaud, Sarah Janvier-Badosa, Kevin Beck, Livio De Luca, and Muzahim Al-Mukhtar. 2012. 3D Information System for the Digital Documentation and the Monitoring of Stone Alteration. In *Progress in Cultural Heritage Preservation*, Marinos Ioannides, Dieter Fritsch, Johanna Leissner, Rob Davies, Fabio Remondino, and Rossella Caffo (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 330–339.

[25] Ivan Ustyuzhaninov, Claudio Michaelis, Wieland Brendel, and Matthias Bethge. 2018. One-shot texture segmentation. *arXiv preprint arXiv:1807.02654* (2018).