MO.SE.: MOSAIC IMAGE SEGMENTATION BASED ON DEEP CASCADING LEARNING

MO.SE.: SEGMENTACIÓN DE MOSAICO DE IMÁGENES BASADO EN APRENDIZAJE PROFUNDO EN CASCADA

Andrea Felicetti[a] , Marina Paolanti[a] , Primo Zingaretti[a] , Roberto Pierdicca[b,*] ,
Eva Savina Malinverni[b]

[a] Dipartimento di Ingegneria dell'Informazione, Università Politecnica delle Marche, via Brecce Bianche 1, 60131 Ancona.
a.felicetti@pm.univpm.it; m.paolanti@univpm.it, p.zingaretti@univpm.it

[b] Dipartimento di Ingegneria Civile, Edile e dell'Architettura, Università Politecnica delle Marche, via Brecce Bianche 1, 60131 Ancona.
r.pierdicca@pm.univpm.it, e.s.malinverni@univpm.it

Highlights:

- A Mo.Se. (Mosaic Segmentation) algorithm is described with the purpose to perform robust image segmentation to automatically detect tesserae in ancient mosaics.

- This research aims to overcome manual and time-consuming procedure of tesserae segmentation by proposing an approach that uses deep learning and image processing techniques, obtaining a digital replica of a mosaic.

- Extensive experiments show that the proposed framework outperforms state-of-the-art methods with higher accuracy, even compared with publicly available datasets.

Abstract:

Mosaic is an ancient type of art used to create decorative images or patterns combining small components. A digital version of a mosaic can be useful for archaeologists, scholars and restorers who are interested in studying, comparing and preserving mosaics. Nowadays, archaeologists base their studies mainly on manual operation and visual observation that, although still fundamental, should be supported by an automatized procedure of information extraction. In this context, this research explains improvements which can change the manual and time-consuming procedure of mosaic tesserae drawing. More specifically, this paper analyses the advantages of using Mo.Se. (Mosaic Segmentation), an algorithm that exploits deep learning and image segmentation techniques; the methodology combines U-Net 3 Network with the Watershed algorithm. The final purpose is to define a workflow which establishes the steps to perform a robust segmentation and obtain a digital (vector) representation of a mosaic. The detailed approach is presented, and theoretical justifications are provided, building various connections with other models, thus making the workflow both theoretically valuable and practically scalable for medium or large datasets. The automatic segmentation process was tested with the high-resolution orthoimage of an ancient mosaic by following a close-range photogrammetry procedure. Our approach has been tested in the pavement of St. Stephen's Church in Umm ar-Rasas, a Jordan archaeological site, located 30 km southeast of the city of Madaba (Jordan). Experimental results show that this generalized framework yields good performances, obtaining higher accuracy compared with other state-of-the-art approaches. Mo.Se. has been validated using publicly available datasets as a benchmark, demonstrating that the combination of learning-based methods with procedural ones enhances segmentation performance in terms of overall accuracy, which is almost 10% higher. This study's ambitious aim is to provide archaeologists with a tool which accelerates their work of automatically extracting ancient geometric mosaics.

Keywords: cultural heritage; mosaic; deep learning; image segmentation; digitization

Resumen:

El mosaico es un tipo de arte antiguo utilizado para crear imágenes decorativas o patrones de pequeños componentes. Una versión digital de un mosaico puede ser útil a los arqueólogos, estudiosos y restauradores que están interesados en el estudio, la comparación y la preservación de los mosaicos. Hoy en día, los arqueólogos basan sus estudios principalmente en la operación manual y la observación visual que, aunque sigue siendo fundamental, debe ser apoyada con la ayuda de un procedimiento automatizado de extracción de la información. En este contexto, esta investigación tiene la intención de superar el procedimiento manual y lento del dibujo de teselas en mosaico proponiendo Mo.Se. (Mosaic Segmentation), un algoritmo que explota técnicas de aprendizaje profundo y segmentación de imagen; específicamente, la metodología combina la red U-Net 3 con el algoritmo Watershed. El propósito final es definir un flujo de trabajo que establezca los pasos para realizar una segmentación robusta y obtener una representación digital (vectorial) de un mosaico. Se presenta el procedimiento detallado y se proporcionan justificaciones teóricas, construyendo varias

---

*Corresponding author: Roberto Pierdicca, r.pierdicca@univpm.it

conexiones con otros modelos, haciendo que el flujo de trabajo sea teóricamente valioso y prácticamente escalable en conjuntos de datos medianos o grandes. El proceso de segmentación automática se probó con la ortoimagen de alta resolución de un mosaico antiguo, siguiendo un procedimiento de fotogrametría de objeto cercano. Nuestro enfoque se ha probado en el pavimento de la Iglesia de San Esteban en Umm ar-Rasas, un sitio arqueológico de Jordania, ubicado a 30 km al sureste de la ciudad de Madaba (Jordania). Los resultados experimentales muestran que este marco generalizado produce buenos rendimientos, obteniendo una mayor precisión en comparación con otros enfoques de vanguardia. Mo.Se. se ha validado utilizando conjuntos de datos disponibles públicamente como punto de referencia, lo que demuestra que la combinación de métodos basados en el aprendizaje con métodos procedimentales mejora el rendimiento de la segmentación en casi un 10% en términos de exactitud en general. El ambicioso objetivo de este estudio es proporcionar a los arqueólogos una herramienta que acelere su trabajo de extracción automática de mosaicos geométricos antiguos.

Palabras clave: patrimonio cultural; mosaico; aprendizaje profundo; segmentación de imagen; digitalización

1. Introduction

In the current scenario of Cultural Heritage (CH), the digitization gained paramount importance for the documentation and interpretation of cultural artefacts (Pierdicca, Frontoni, Malinverni, Colosi, & Orazi, 2016; Pierdicca et al., 2015). This process is true even for mosaics. Mosaic is an ancient type of art used to create decorative images or patterns of small components (Battiato, Di Blasi, Farinella, & Gallo, 2007).

Dealing with mosaics is challenging and fascinating, given their uniqueness with respect to other figurative arts. Their creation was laborious, expensive and very time consuming but, conversely, the result more persistent in time. For this reason, it can be found in areas where painting could not be feasible, such as floors. During the IV century BC, emerges what the mosaic technique par excellence: the *opus tessellatum*, the mosaic with tesserae. Afterwards, along with pebbles are commonly employed glass, ceramic and stone tesserae and modernly, any small component with traditional materials can be used: glass or ceramic cast or cut into tiles, also plastic, beads, buttons, bottle caps, pearls, and more (Benyoussef & Derrode, 2011).

Given the above and due to the complexity of the subject, owning a "digital replica" of mosaics is mandatory, since it represents the starting point for the preservation and the valorization at a worldwide scale. Indeed, archaeologists, scholars and restorers can exploit digital technologies for studying, comparing and preserving the mosaics. Moreover, for restorers, a digital model of a mosaic could become an essential professional tool. Researchers are thus supplied with highly reliable models, so to have new reading keys to study monuments (Fontanella et al., 2019). Furthermore, it serves as a record of their state-of-conservation and, at the same time, as a method by which they can be preserved (Bourke, 2014). Different issues concerning their exploitation are still cause of discussion. The aspects under investigations involve data reliability and the needs of dissemination/interaction or the use of digital models as bases for traditional drawings (Cipriani & Fantini, 2017).

Up to now, actions like segmentation, information extracting, and labelling tesserae are made manually. Indeed, this procedure is intensive labour for archaeologists, and it is performed manually by experts. To overcome the above limitations, this paper outlines a novel approach using deep learning and image processing methods for the automatic tesserae segmentation.

Besides extending the approach and the analysis presented in (Felicetti et al. 2018), this work attempts to define Mo.Se. (Mosaic Segmentation) as an algorithm which set closely related steps for: i) performing a robust image segmentation to automatically detect tesserae in ancient mosaic, ii) managing dedicated information within a geodatabase for understanding the evolution of the iconographic repertoire, as described in Malinverni, Pierdicca, Di Stefano, Gabrielli, & Albiero, (2019), iii) applying the workflow on the pavement of St. Stephen's Church in Umm ar-Rasas, a Jordan archaeological site, located 30 km southeast of the city of Madaba, in the northern part of Wadi Mugi; iv) validating the experiments upon the state-of-the-art (SoA) benchmark dataset, proposed in Fenu, Medvet, Panfilo, & Pellegrino (2020); v) providing to the archaeologists a tool for automatic extraction of geometric of ancient mosaics, facilitating their daily work.

The paper is organized as follows: Section 2 provides a description of the approaches that were adopted for the segmentation task. Section 3 gives details on the proposed workflow, which is the main core of our work. In Section 4, an extensive comparative evaluation of our approach with respect to the state-of-the-art is offered, as well as a detailed analysis of each component of our approach. Finally, in Section 5, conclusions and discussion about future directions for this field of research are drawn.

2. Related works

Despite computer science and image processing are largely used in the CH domain, few pieces of research involving such disciplines for mosaic conservation, restoration or cataloguing are reported in the literature (Fenu, Jain, Medvet, Pellegrino, & Namer, 2015). As well, the contributions of artificial intelligence in this domain seems still neglected (Bordoni & Mele, 2016).

To acquaint the reader about the latest research trends, some examples are reported, together with the baseline that we used for our study.

The largest number of related works faces the issue of shapes interpretation and pattern recognition, mainly for cataloguing purposes, which are usually based on image processing techniques.

In Zitová, Flusser, & Šroubek (2004), the authors present an application of digital image processing techniques in medieval mosaic conservation. Their case study was *The Last Judgment mosaic*, located on the wall of the St. Vitus Cathedral in Prague, in the Czech Republic. They have compared the historical photograph of the mosaic from the 19th century, with the current photograph to detect mutual differences. At first, they have pre-processed the images to increase their quality (noise reduction,

deblurring). Then, they have removed the geometrical differences between images with image registration techniques, mutual information and feature point correspondence. Lastly, they have determined the differences between the historical and the current photographs.

In Zarghili, Gadi, Benslimane, & Bouatouch (2001) and Zarghili, Kharroubi, & Benslimane (2008), the authors have paid attention to Islamic mosaics, which have particularities in the periodicity and symmetry of tessera patterns, by proposing a method to index an Arabo-Moresque decor database which is not based on symmetry. In particular, they have used a supervised mosaicking technique to identify the main geometric information (connected set of polygonal shapes, namely "spine") of a pattern. By adopting Fourier shape descriptors for allowing retrieving of images even under translation, rotation and scale they have described the spine. However, this method cannot be automatized and does not allow the pattern classification according to any criteria (Zarghili et al., 2001; Zarghili et al., 2008).

In line with the above-mentioned studies over Islamic art, other works are noteworthy. Djibril, Thami, Benslimane, & Daoudi (2005) for instance, designed a system to index Arabo-Moresque mosaic images with symmetry and auto-similarity of motifs. It was based on the fractal dimension. The automatic segmentation was performed using colour information. The original motif was decomposed into a set of basic shape after the classification. The shapes contours are characterized by their fractal dimension which provides a significant measure of the geometric structure of the tessera pattern.

The study of Islamic geometrical patterns, and all periodic patterns such as those encountered in textile patterns or wallpapers, was approached even by the symmetry group theory. In Djibril & Thami (2008), as first step authors have classified patterns in three classes: pattern created after translation along one dimension, patterns with translational symmetries in two independent directions, and "rosettes", i.e. patterns that start at a central point and grow radially outward. For each pattern, the symmetry group and the primary region have been extracted. After that, they characterize the primary region by a colour histogram and build the feature vector.

Lastly, in Gil, Gomis, & Pérez (2009), the authors describe image processing techniques to restore mosaic patterns. They have developed an image analysis tool to obtain information about design patterns which are employed to recover tesserae or missing motifs. A great obstacle was in proposing a method that results robust to the discrepancies between equal object shapes. Once the symmetry has been recovered, it allows the virtual reconstruction by inpainting methods and physical restoration of damaged parts of mosaics.

Being the creation of information systems essential for the management of information, in the literature, some works that face this matter.

A Content-Based Image Retrieval (CBIR) system to index and catalogue Roman mosaic images have been proposed in (M'hedhbi, Mezhoud, M'hiri, & Ghorbel, 2006; Maghrebi, Baccour, Khabou, & Alimi, 2007; Maghrebi, Borchani, Khabou, & Alimi, 2007). This system includes an object extraction from a complex mosaic scene by using unsupervised statistical segmentation and an invariant description of semantic objects using the analytical Fourier-Mellin transform. An index created from the invariant descriptors and an appropriate metric (Hausdorff and Euclidean) gives the similarity between querying mosaic and the database.

In Maghrebi, Baccour, Khabou, & Alimi (2007) and Maghrebi, Ammar, Alimi, & Khabou (2013), the authors describe a CBIR which is a general system to index and retrieve by the content historic document images using a mouse or a pen, a user drawing query, so that a query of pertinent shapes from the database begins. A measure of Fuzzy similarity is used to compare entries in the database. An XML database is integrated into the system, and experiments on huge databases of some Tunisian museums and the National Library of Tunisia are listed.

Another important task is the automatic detection of tesserae. To the best of our knowledge, a single mosaic-oriented segmentation algorithm has been proposed in the literature (Youssef & Derrode, 2008), that is based on the well-known watershed algorithm (Vincent & Soille, 1991) and some mosaic-specific pre-processing steps.

The solution adopted by the authors is based on grey-level morphology which is suitable to the tiling organisation of mosaics. Furthermore, they have presented a Watershed Transformation (WT) approach. The quality of the extraction depends on the image acquisition mode.

Taking into account the technique implemented to perform image mosaic segmentation, some recent studies use an approach based on deep learning. In the work of (Fenu et al., 2020) the authors exploit the U-Net network (Ronneberger, Fischer, & Brox, 2015), a convolutional neural network, to perform the image mosaic segmentation so that each segmented region precisely corresponds to a tessera of the mosaic and processes the image at the pixel level. They use this approach since they retained that was most useful in different segmentation tasks (Çiçek, Abdulkadir, Lienkamp, Brox, & Ronneberger, 2016; Kohl et al., 2018; Falk et al., 2019). They compare the performances of the proposed method with other segmentation approaches dedicated to mosaic tesserae, that as stated by the authors themselves, are very scarce.

The work presented in Bartoli, Fenu, Medvet, Pellegrino, & Timeus (2016) has the same purpose where, however, deformable models are used to overlap the mosaic and adapt to the actual shape of the tesserae. To optimize the deformable forms, they use genetic algorithms.

3. Materials and methods

The experimental approach presented in (Felicetti et al., 2018) has been improved by developing Mo.Se., an algorithm specifically conceived for the segmentation of ancient mosaics. In this section, the case study, the training dataset and the process are described, together with the benchmark datasets used for the evaluation.

3.1. Case study: The Byzantine mosaics in the Church of St. Stephen (Jordan)

The mosaic chosen for our study is located in the Church of St. Stephen, in Jordan. The pavement is decorated with an ancient Byzantine mosaic containing inscriptions, portraits of the donors, geometric and vegetal motifs, representation of cities and scenes from the world of pastoralism, agriculture, hunting, combined with marine motifs and Nilotic, arranged in a precise decorative

structure. Its distribution is related to the internal organization of the building and linked to the liturgical destination of the areas. Through non-verbal language, the scenes of this mosaic had the task to transmit eschatological, economic, religious and even historical teachings, like a sort of storytelling for the communities leaving these countries.

Data were acquired through different surveying techniques, such as laser scanner and digital photogrammetry. For implementing the so-called Multi-View Stereo method, a huge amount of shots (about 4000 in total) was captured; as results, we obtained the best radiometric quality of the surfaces and geometrical detail of the floor. This acquisition was carried out with a reflex camera Canon EOS 5D Mark II with a full-frame sensor (21.1 MP) and a 28 mm camera lens. The camera was arranged on the top of a pole and constantly held at a height of 170 cm. The scans were performed with a TLS instrument, the Faro Focus3D 120. Combining with the point cloud model, made by the close-range photogrammetry, it was possible to obtain a 3D metric model of the floor. The integration of the two survey techniques, georeferred in a local system, has generated an orthomosaic in 1:1 scale of the floor surface, characterized by the descriptive accuracy of colours, and details of the surveyed surfaces. The Digital Elevation Model (DEM) with an accurate representation of the geometric deformations and a corresponding draped orthoimage is shown in Figure 1. The orthoimage has reached a Ground Sampling Distance (GSD) of 0.21 mm.

3.2. Mosaic datasets

In this Section, the datasets used to train and assess the experiments with Mo.Se. approach, are described. Following the work described in (Malinverni et al., 2019),

the pavement of the Byzantine mosaics of the Church has been manually segmented to create a dataset, partially used as training data and partially as ground truth (gt); more in deep, for the training set 50 out of 100 samples are selected, about 1.25 m^2 of gt and the remaining 4.25 m^2 for testing. Furthermore, in this paper to validate and generalise Mo.Se., open access mosaic datasets have been used. The data collection and annotation are presented in the following subsections, as well as the SoA datasets used for the testing.

3.2.1. Training and ground-truth dataset

Since our approach is learning-based, the first step was to label the training data, used as the ground truth of the model. This step was manually made by expert archaeologists. For the annotation phase, the QGIS software has been exploited. The orthoimage of the georeferenced mosaic has been loaded and vectorial layers manually performed on it.

First of all, the whole image of mosaic has been subdivided into 18 panels. By panels is meant portions of mosaic that represent individual scenes, sometimes separated by frames or tesserae of the same colour set in periodic patterns that form straight lines, waves, etc. A shapefile "panel", made of polygons (multipolygons), has been created to map the panels of the mosaic, taking into account their contents. Figure 1 shows the panels superimposed on the orthoimage of the mosaic. Each multipolygon has been univocally identified by a letter of the alphabet.

A second annotation procedure elaborates the single tesserae. For this purpose, a shapefile "tesserae" is made of polygons which approximately identify the contour of each tessera. Being a very expensive and time-consuming task, the single tesserae annotation has not
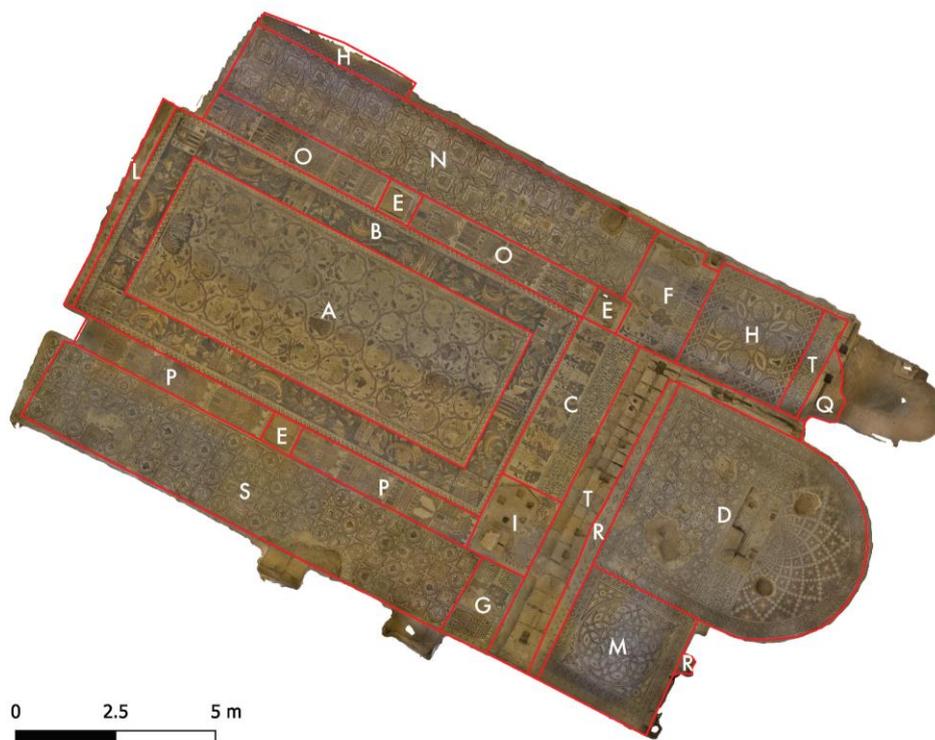
Figure 1: Complete orthoimage created for the case study, inside the Byzantine Church of St. Stephen. The red rectangles highlight different panels identified by archaeologist, taking into account the historical subject. Total surface 225 m^2, gt surface 5.5 m^2 (54774 tesserae).

been made on the whole mosaic. Only 100 squares of size 672x672 pixels scattered between the B, C, G, F, S panels of the mosaic have been selected. The squares are selected to contain an equal and representative part of the whole mosaic. The squares can be overlapped on each other by 224x224 pixels. The single tesserae annotation has been made inside these 100 squares. However, a tesserae annotation previously made in other areas of B, C, G, F, S panels has been added to the gt. This last annotation is only used in the testing phase.

3.2.2. SoA mosaics dataset

The dataset described in (Fenu et al., 2015; Fenu et al., 2020; Bartoli et al., 2016) has been chosen for Mo.Se. evaluation and generalization. This dataset has been collected by considering five mosaic images, representing five mosaics with a different style and age. In particular, the images depict the mosaic forming the floor of the Basilica di Santa Maria Assunta, in Aquileia (Udine, Italy); mosaic at the Early Christian Museum (Museo paleocristiano di Monastero) in Aquileia; flower a small contemporary mosaic, built by an Italian amateur as an essay for a course of ancient mosaic technique; University, a portion of a mosaic of the floor in the building of a campus, aged 1938.

3.3. Mo.Se.: algorithm for automatic tesserae image segmentation

Automatic tesserae image segmentation is performed by implementing Mo.Se., which performs the steps defined in the algorithm. In particular, it comprises the following phases, depicted in Figure 2:

a) High Definition (HD) acquisition of Red-Green-Blue (RGB) orthoimage of the mosaic pavement of St. Stephen's Church (Malinverni et al., 2019);

b) Tesserae image segmentation using U-Net 3 Deep Neural Network;

c) Hierarchical Watershed Algorithm;

d) Refining of segmentation;

e) Raster to vector conversion of segmentation for approach performance comparison with SoA datasets (Fenu et al., 2020).

Algorithm 1 (Tessera image segmentation) depicts the steps performed by Mo.Se. algorithm for the automatic image segmentation of mosaics.

Firstly, U-Net3 network (Liciotti, Paolanti, Pietrini, Frontoni, & Zingaretti, 2018) has been chosen for this task. The model is trained from zero, starting from a random initialization of weights. During the training phase,

the input RGB images have a dimension of 224x224 pixel. Three normalization strategies have been evaluated:

1) normalization on the single input,

2) normalization on the batch,

3) normalization on the training set.

The normalization on the training set is useful for the recomposition, to avoid problems of discontinuity along the stitching edges.

Algorithm 1: Tessera Segmentation.

```
Require: orthophoto RGB of mosaich, segmentation NN trained
 1: clips ← clipping (windowsDimension, overlap, stride= windowsDimension - overlap)
 2: for clip ← clips do
 3:     neural network prediction of clip (windowsDimension)
 4:     cutting overlapped borders of clip (overlap)
 5:     merging in neural network prediction of mosaic orthophoto
 6: end for
 7: if normalization over clip used then
 8:     sewing lines filtering (L,k)
 9: end if
10: image inversion
11: for shallowMinimaThreshold ← (127, 63, 31, 15, 7, 3, 1, 0) do
12:     regions ← watershed(shallowMinimaThreshold)
13:     for region ← regions do
14:         if regionarea ≥ minimumAreaToBeSubdivided then
15:             goto next level of watershed
16:         else
17:             tesseraregion ← region
18:         end if
19:     end for
20: end for
21: for tesseraregion ← tesseraregions do
22:     histogram of tessera region
23:     threshold ← half value between 0 and knee of histogram
24:     tessera ← tresholding of tessera region (threshold)
25: end for
```

We have chosen to adopt a batch normalization approach during the training for achieving better results in the segmentation. The target (binary image) was mapped with 0 and 1 values, corresponding to black and white respectively. The model processes the input image generating in the output a prediction image compared with the target image. The error between prediction and target, evaluated according to the loss function in Eq. (1) is used to modify the model weights. The goal of training is to minimize the loss function, while Adam optimization algorithm (Kingma & Ba, 2014) is an extension to stochastic gradient descent that has recently seen broader adoption for deep learning applications in computer vision and natural language processing.

$$Loss = 1 - \frac{\sum pred * gt}{\sum pred^2 + gt^2} \tag{1}$$

A mini-batch training strategy with a batch size of 16 has been adopted. This arises from the error evaluation and the consequent modification of the weights occurs on a batch of 16 training pairs. At the same time, the batch size
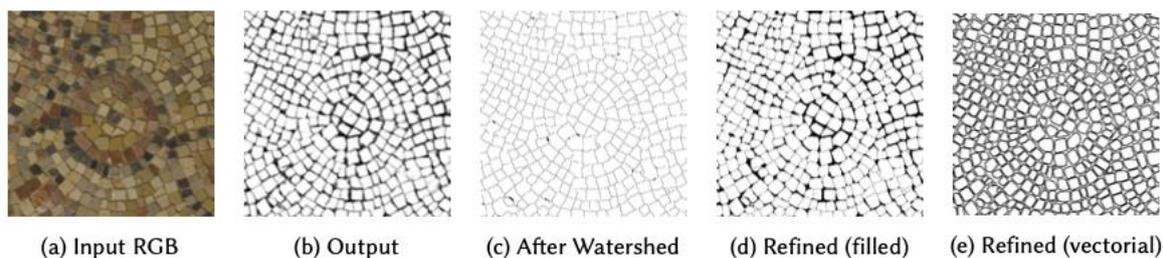

(a) Input RGB   (b) Output   (c) After Watershed   (d) Refined (filled)   (e) Refined (vectorial)

Figure 2: Step by step representation of Mo.Se. algoritm for Image Segmentation.

set to 16 guarantees stability and rapidity of convergence. The training procedure has been performed in 3 periods each of 200 epochs. For each period a different learning rate has been set for the first 0.01, for the second 0.001 and for the third 0.0001. After the training phase, the neural network can segment any RGB image of the mosaic. Furthermore, thanks to the spatial invariance of the convolutional neural networks (except for the edges of the window) it is possible to use sliding windows to process images larger than the input of the network (Bonfigli et al., 2018). Moreover, since the U-Net is fully convolutional, the dimension of the input can be changed without altering the internal parameters. During the cropping of the mosaic image, for each square, the coordinates are preserved in order to recompose the original image. For each input square, the segmentation procedure was executed. From the segmented squares, the image of the mosaic segmentation at the output of the neural network is recomposed. The squares recomposition creates some artefacts (discontinuities) along the stitching edges. This effect is due to the poor segmentation accuracy of the network at the edge of the image and to normalization with parameters (average value and standard deviation) that vary between one crop and another. To solve this problem during the crop of the RGB image, an overlap of the squares of 256 pixels per edge is guaranteed.

During stitching, the segmented square is trimmed by 128 pixels per edge (those with low accuracy) and joined to the adjacent square, which is also trimmed by 128 pixels per edge.

The trimming of 128 pixels per edge ensures the absence of discontinuity. A smaller value could maintain discontinuity while a larger value would weigh on the computational cost. Using the normalization strategy on the single input, the trimming of the segmented squares is not sufficient to eliminate the artefacts. To mitigate these discontinuities, an average value filtering is performed along the stitching edges.

To separate the individual tesserae among them, the output image from the neural network is binarized using an improved version of the Watershed Algorithm (Vincent & Soille, 1991) which is named Hierarchical.

The Watershed Algorithm is based on region growing method; the greyscale image in output from the neural network is reversed and mapped into a space of range values from 0 to 255. In this 3D representation (x, y, values), the predicted tesserae can be associated with valleys while the interspaces to ridges that separate the valleys. Below a predefined height (value) wells are dug and each well corresponds to a basin. The watershed consists of flooding the selected basins up to the ridges (Beucher, & Lantuéjoul, 1979) where each basin contains the predicted tessera.

The Hierarchical Watershed Algorithm is applied by defining a threshold with a "minimum area to be subdivided" only to regions with the higher area than this threshold. The Hierarchical Watershed Algorithm requires the setting of the threshold (suppressing shallow minimum) to extract the seeds of the predicted tesserae. The latter is increased to fill the basin delimited by the interspaces. Using a too high threshold, many of the individual tesserae remain aggregated in regions of large area (under-segmentation). Using a too low threshold, many of the individual tesserae are subdivided into small area regions (over-segmentation). The "shallow minima" thresholds used during the experimental phase are: 127, 63, 31, 15, 7, 3, 1, 0. The dependence of watershed from the threshold is studied during preliminary analysis. The choice of thresholds in the scaling of "powers of two minuses one" ensures a correct variation in the number of tesserae extracted between one level and another, able to perform the hierarchical watershed with good efficiency and accuracy. Optimization in the choice of thresholds and the number of levels are not executed.

Starting from the higher shallow minima threshold, the area of the regions obtained is measured. At the next level, the watershed algorithm with a lower shallow minimum threshold is applied only to regions that have an area greater than the "minimum area to be subdivided" threshold. In the same way, the watershed algorithm is applied at all levels. Moreover, a rule-based on the eccentricity has been added to subdivide the region under the threshold. This is done because the tesserae are square-shaped with an eccentricity under the threshold.

At each level, the regions of a small area that cannot be considered tesserae and are associated with interspace and filled with black. Many of these regions correspond to areas where the interspace between the tesserae is thicker than the average, recurrent in the intersections between interspaces, and where there are considerably damaged tesserae.

Regions with an area above an established threshold, whose average value of the pixels in the segmentation image at the output of the neural network is less than 0.05 are associated with interspace and filled with black. The same rule is not applied to small area regions because it is more likely that it is a small tessera (between large interspaces). The watershed algorithm to the next level is not applied to the regions associated with interspace. Regions that do not respect these rules are classified as non-mosaic.

To evaluate the improvements obtained using the Hierarchical Watershed Algorithm, 3 classes are defined: "small", "medium" and "big". They identify predicted tesserae whose dimension presupposes that they are fragmented, corrected, and joined to others. From a statistical analysis of the area of the gt tesserae (Fig. 3), the average area is calculated, and the thresholds are defined to classify the predicted tesserae. The gt contains 54774 tesserae. The smaller area tessera measures 4.42 mm^2 while the larger area tessera measures 341.15 mm^2. The histogram was calculated in the range from 0 to 350 mm^2 using 70 bin, every 5 mm^2 in size. The average area of the tesserae is 73.5mm^2. The histogram is calculated in the range of 0 to 1 using 50 bins. The average eccentricity of the tesserae is 0.66 and is approximately equal to the eccentricity of an ellipse with a larger and smaller diameter in a 4:3 ratio, so the eccentricity is $\sqrt{7}/4$.

Three thresholds are defined in the areas (Table 1). The "minimum area", chosen equal to the area of the smallest gt tessera, is defined to exclude from the evaluation all the regions of smaller area (the region of smaller area than this threshold does not contain tesserae but is a prediction error). The "lower area", chosen equal to one half of the average area of the gt tesserae, is defined to separate the small and medium classes. The "upper area", chosen equal to 3 times the average area of the gt tesserae, is defined to separate the medium and big
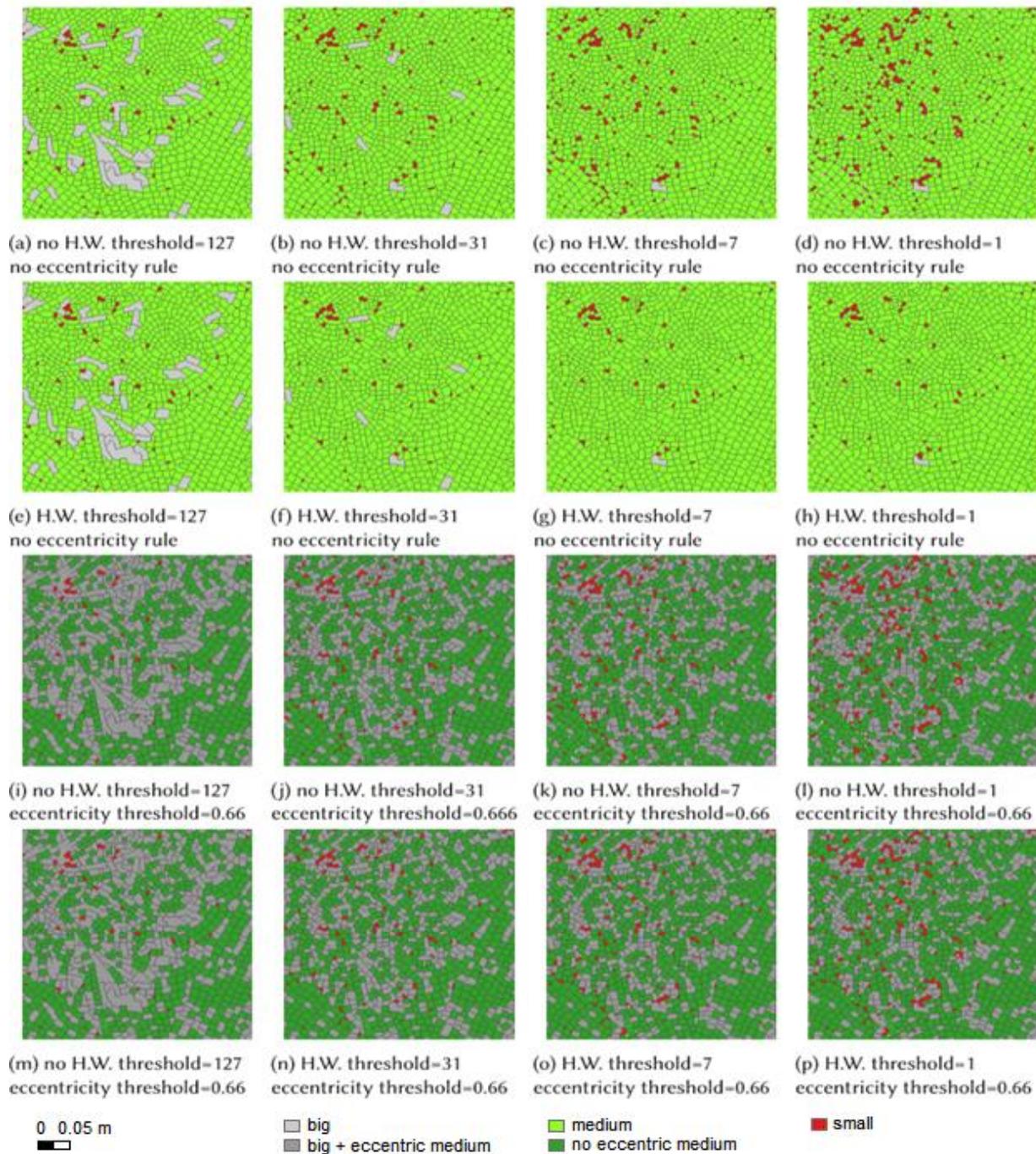
Figure 3: Hierarchical Watershed Algorithm with different eccentricity thresholds. The figures from a) to h) are the results of the algorithm without the eccentricity rule. These figures show the predicted tesserae "big", "medium", "small" coloured respectively in grey, green and red. The figures from i) to p) are the results of the algorithm with the eccentricity rule. These figures show the aforementioned tesserae "big + eccentric medium", "no eccentric medium", "small" coloured respectively in grey, green and red.

classes. The upper area threshold is equivalent to the minimum area to be subdivided threshold used in the Hierarchical Watershed Algorithm. Considering these thresholds there are 1791 gt tesserae less than the lower area threshold and 157 gt tesserae greater than the upper area threshold. They are respectively equal to 3.2% and 0.3% of the total gt tesserae.

Table 1 reports the thresholds set for the evaluation after the statistics on the tesserae.

Table 1: Thresholds defined based on the tessera area values.

Thresholds	Values
Minimum area	4.42 mm^2
Lower area	36.75 mm^2
Upper area	220.5 mm^2
Eccentricity	$\sqrt{7}/4$

The image processed by the Hierarchical Watershed Algorithm appears as a binary image in which the white regions associated with the predicted tesserae are separated by a 1-pixel thick black outline. This segmentation does not take into account the thickness of the interspace. For this, the image is taken at the output of the neural network and a Refining phase is applied to the image. Each region separated from the Hierarchical Watershed Algorithm (basin of the predicted tessera) is used to select the corresponding region on the output image of the neural network.

The histogram of this region is evaluated, and the threshold calculated as the half between zero and the knee of the highest peak. The calculated threshold is used to binarize the respective region. After, an "and" is made between the resulting image and the image after the watershed to ensure the closure of the regions identified by the watershed.

3.4. Mo.Se. algorithm performance evaluation metrics

The metrics used to evaluate the performance of the segmentation at the output of the neural network processing are Intersection over Union (*IoU*) (Eq. 2) or Jaccard Index and *Dice* corresponding to the *F_1 score* (Eq. 5). *Precision* (Eq. 3) and *Recall* (Eq. 4) were also assessed.

Accuracy has not been used, as it is only a global metric, not significant in case of very unbalanced classes. In fact, in our case, the area under the tesserae is greater than the area under the interspaces.

$$IoU = \frac{\sum pred \cap gt}{\sum pred \cup gt} \tag{2}$$

$$Precision = \frac{\sum pred \cap gt}{\sum pred} \tag{3}$$

$$Recall = \frac{\sum pred \cap gt}{\sum gt} \tag{4}$$

$$Dice = F_1 score = 2\frac{\sum pred \cap gt}{\sum pred + gt} \tag{5}$$

To evaluate the quality of the segmentation, the refined image is compared with the gt of the tesserae. Since gt is a vector, the refined image is converted to vector.

To select the predicted tesserae in the area where gt is present, a spatial join is performed between the predicted vector and gt. Then the individual gt tesserae are considered and how many predicted tesserae are overlapped with the gt tessera.

Since there is no unequivocal correspondence between the predicted tesserae and gt superimposed on each other, the "winner take all" (WTA) strategy is adopted to obtain unequivocal correspondence.

For each tessera of gt the areas intersecting with the predicted tesserae are considered, and the predicted tessera with the greatest intersection area is associated with the tessera of gt. This correspondence is not bi-unequivocal since a predicted tessera can be associated with more tesserae among those of gt.

Then a percentage of overlap between the areas of a predicted tessera and the corresponding one of the gt is defined as a threshold "overlap constrain".

The number of matches between the predicted tesserae whose intersection area with the gt tesserae exceeds the threshold is counted.

Accuracy is defined as the relationship between the number of matches and the number of gt tesserae.

Mathematically the binary images of the segmentations are considered: ground truth (gt) and predicted (pred). Each is represented as the sum of as many instance images as there are instances (tesserae) (Eq. 6).

$$gt = \sum_{i=1}^{N} gt_i$$

$$pred = \sum_{i=1}^{M} pred_i \tag{6}$$

gt_i and $pred_i$ are instance images. *N* and *M* indicate the number of instances of gt and predicted segmentations respectively.

For both segmentations, it is assumed that several instance images have not overlapping tesserae.

In other words:

$\sum gt_i \cap gt_j = 0$ e $\sum pred_i \cap pred_j = 0 \ \forall \ i{\neq}j$.

$$Match_i(ovC) = \begin{cases} 1 \ se \ \frac{int_{ij}}{gt_i} \geq ovC \wedge \frac{int_{ij}}{pred_i} \geq ovC \\ 0 \ otherwise \end{cases} \tag{7}$$

$\forall i \in [1, N]$

$$Accuracy(ovC) = \frac{1}{N}\sum_{i=1}^{N} Match_i(ovC) \tag{8}$$

Abbreviation *ovC* means *ovConstrain.*

The metrics used in the comparison with the SoA are the same as those used in (Fenu et al., 2015; Fenu et al., 2020). Defined as:

$$Precision = \frac{1}{N}\sum_{i=1}^{N} max_{j=1}^{M} \frac{\sum pred_j \cap gt_i}{\sum pred_j} \tag{9}$$

$$Recall = \frac{1}{N}\sum_{i=1}^{N} max_{j=1}^{M} \frac{\sum pred_j \cap gt_i}{\sum gt_i} \tag{10}$$

$$F_1 score = 2\frac{PrecisionRecall}{Precision + Recall} \tag{11}$$

$$Cnt = \frac{|N - M|}{N} \tag{12}$$

4. Results

In this Section, the first part presents the results of each phase performed by Mo.Se. Algorithm. The second one includes the comparison among different SoA datasets.

4.1. Mo.Se. validation by parametric indices

Table 2 presents the results of the segmentation in the output of the neural network using the UNET-3. Metrics

are determined after the binarization of the neural network output with a 0.5 threshold.

Table 2: Results of segmentation at the output of the neural network.

IoU	Precision	Recall	Dice
0.79	0.92	0.94	0.93

For the evaluation of the proposed Hierarchical Watershed Algorithm, we compare the results with the use of the eccentricity. Figure 4 depicts the results of this comparison with different eccentricity thresholds. Instead, Figure 5a and 5b graphically report this comparison.

(a)   (b)
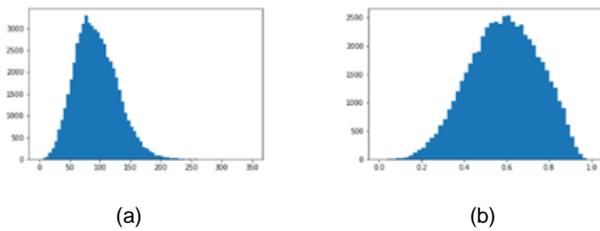
Figure 4: Results of statistics on the area and eccentricity of gt tesserae. In x-coordinates the values of area (a) and eccentricity (b) of tesserae. In y-coordinates the number of tesserae belonging to a corresponding interval.
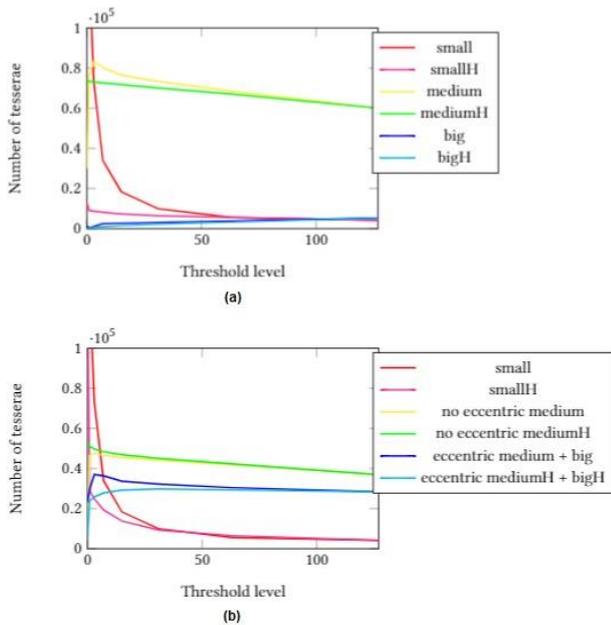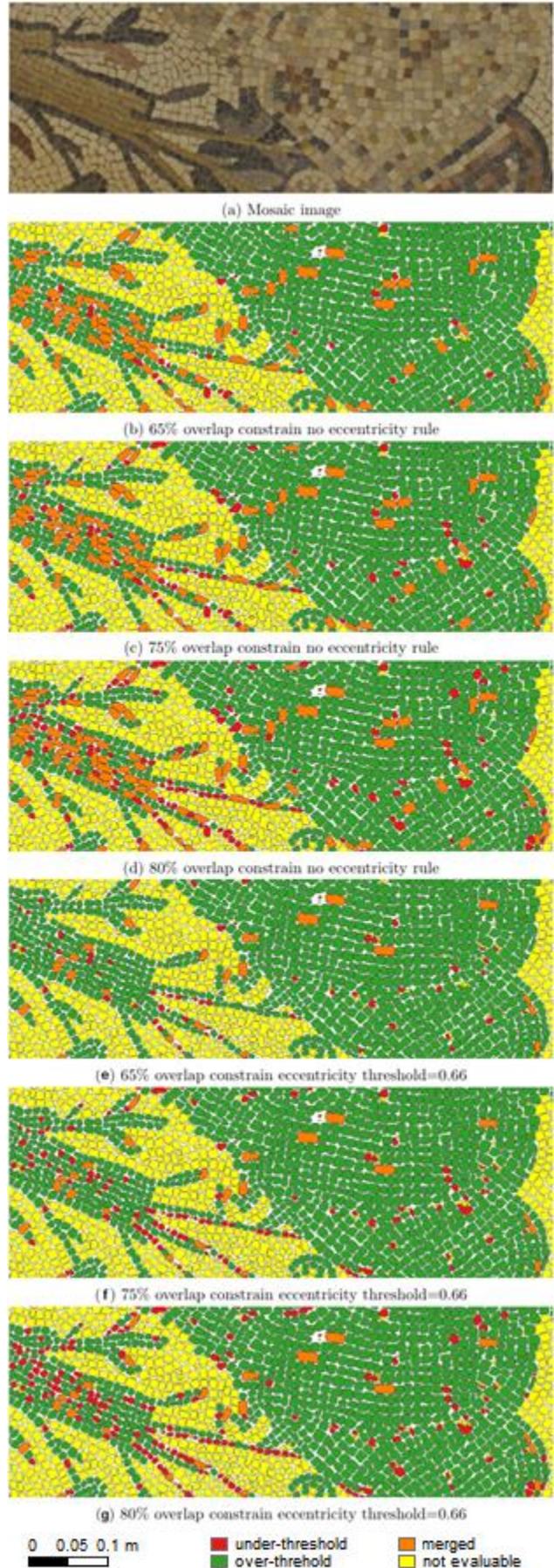

Figure 5: (a) Representation of Hierarchical Watershed Algorithm and (b) reports the results of Hierarchical Watershed Algorithm using eccentricity rule.

It can be observed that applying Hierarchical Watershed Algorithm, at low thresholds, the number of small regions due to over-segmentation is reduced.

Consequently, many medium regions are recovered. However, an identical number of big regions remains.

Figure 6 represents the results obtained after the Refining phase. From this Figure that represents a part of the results in Tables 3 and 4, it can be observed that the introduction of the eccentricity rule considerably reduces

(a) Mosaic image

(b) 65% overlap constrain no eccentricity rule

(c) 75% overlap constrain no eccentricity rule

(d) 80% overlap constrain no eccentricity rule

(e) 65% overlap constrain eccentricity threshold=0.66

(f) 75% overlap constrain eccentricity threshold=0.66

(g) 80% overlap constrain eccentricity threshold=0.66

0   0.05 0.1 m   ■ under-threshold   ■ merged
                  ■ over-threshold   ■ not evaluable

Figure 6: Tessera segmentation results. The colours red, green, and orange are respectively the predicted tesserae under-threshold, over-threshold, merged. Yellow tesserae have not the corresponding ground truth so are not evaluable.

the merged tesserae. This ensures the recovery of the highest number of over-threshold tesserae and a moderate increase of under-threshold tesserae. The increase or decrease of the overlap constrain threshold spreads the quantity between over-threshold and under-threshold. If the overlap constrain increases, the number of under-threshold tesserae increases and decreases the number of over-threshold tesserae; on the contrary if the overlap constrain decreases. On the number of merged tesserae does not affect.

It is possible to infer that Mo.Se. algorithm comprises three important phases (U-Net 3, Hierarchical Watershed and Refining) and reaches high accuracy in tesserae segmentation. Tables 3, 4 and 5 show how increasing the required overlapping area, fewer tesserae satisfy this constraint. As a result, accuracy decreases. The same results are shown in Figure 7.

**Table 3**: Results of tessera segmentation.

% OvC	Under-threshold	Over-threshold	Merged	No assigned	Accuracy
51	523	15476	798	19	0.92
65	978	15021	798	19	0.89
75	1896	14103	798	19	0.84
80	3136	12863	798	19	0.76
85	5849	10150	798	19	0.60

Table 4: Results of tessera segmentation without eccentricity rule.

% OvC	Under-threshold	Over-threshold	Merged	No assigned	Accuracy
51	231	14100	2466	19	0.84
65	530	13801	2466	19	0.82
75	1259	13072	2466	19	0.78
80	2336	11995	2466	19	0.71
85	4757	9574	2466	19	0.57

Table 5: Results of tessera segmentation.

Eccentricity	Precision	Recall	F_1score
Y	0.92	0.85	0.89
N	0.87	0.87	0.87
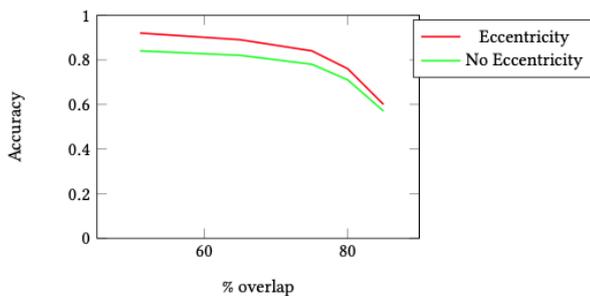
Figure 7: Accuracy over overlap constrain with and without eccentricity.

4.2. Mo.Se. validation by SoA dataset

Our method has been compared to other SoA methods. Following the same test procedure and using the same dataset we compared the results obtained using the methodologies proposed in (Fenu et al., 2020) with the results obtained by implementing our method. Tests were carried out following a leave-one-out strategy. For each method compared, as many tests are performed as there are input-target pairs. For each test, only one pair is used. The remaining pairs are used for training the model. Finally, the results of the test are averaged to obtain the overall accuracy measurements of the method.

Five tests were performed, one for each pair of images. The images have been scaled to the dimensions whereby the average area of the tesserae corresponds to around 2500 (±1500).

In Table 6 are shown the scale factors used to scale the image size.

Given the variety of colours between one image and the other, we have chosen to use our U-Net3 model with a single input channel to process the image of the greyscale mosaic. The training strategy and the set of hyperparameters are those described in Section 3.3.

Table 6: Comparison of dataset image scaling.

Image	Scaling factor
7	0.25
8	1.5
9	1.5
10	0.8
11	1.2

To estimate the improvement introduced by the hierarchical watershed plus the refining, the metrics were evaluated at the output of the neural network and after refining. To evaluate the segmentation of the neural network, the output greyscale image was binarized with a threshold of 0.5.

The comparison of the methods is shown in Table 7. Our segmentation method outperforms other methods in all the tests.

Furthermore, it can be observed that the hierarchical watershed plus refining applied downstream of the neural network improves performance. The same results can be appreciated from the images of the segmentation in Figure 8.

5. Discussion

The results showed above deserve some considerations, in order to outline the pros and cons, besides highlighting limitations and future research directions of the proposed methodology.

The first one concerns the evaluation metrics used to assess and validate Mo.Se.; our main purpose was to evaluate the accuracy of the segmentation phase. For this purpose, two different measures are considered: the capability of the algorithm to discriminate tesserae from one another, and the capability of identifying the correct perimeter. Pixel-based metrics are more general, as they are able just to provide an overall picture of the

Table 7: Comparison results.

a) Results of Fenu et al., 2020

ImageID	Unet				GA				TOS			
	Cnt	Prec	Rec	Fm	Cnt	Prec	Rec	Fm	Cnt	Prec	Rec	Fm
7	0.30	0.65	0.80	0.73	0.03	0.50	0.76	0.60	0.14	0.64	0.87	0.74
8	0.28	0.59	0.71	0.64	0.03	0.42	0.63	0.50	0.54	0.56	0.72	0.63
9	0.21	0.62	0.73	0.67	0.01	0.41	0.66	0.51	0.03	0.53	0.82	0.64
10	0.52	0.65	0.70	0.67	0.07	0.50	0.63	0.56	0.06	0.49	0.68	0.57
11	0.29	0.65	0.72	0.69	0.03	0.46	0.67	0.55	0.90	0.63	0.78	0.70
Avg	0.32	0.64	0.73	0.68	0.03	0.46	0.67	0.54	0.33	0.57	0.77	0.66

b) Our results

ImageID	Unet3				Unet3+Watershed+Refining			
	Cnt	Prec	Rec	Fm	Cnt	Prec	Rec	Fm
7	0.21	0.17	0.92	0.28	0.25	0.75	0.90	0.82
8	0.29	0.06	0.89	0.11	0.19	0.70	0.64	0.68
9	0.31	0.05	0.93	0.09	0.33	0.60	0.87	0.71
10	0.03	0.20	0.78	0.32	0.03	0.80	0.74	0.77
11	0.13	0.25	0.85	0.38	0.14	0.71	0.83	0.77
Avg	0.19	0.15	0.87	0.24	0.19	0.71	0.80	0.75

(a) Unet   (b) GA   (c) TOS

(d) Unet3   (e) Unet3 + Watershed + Refining
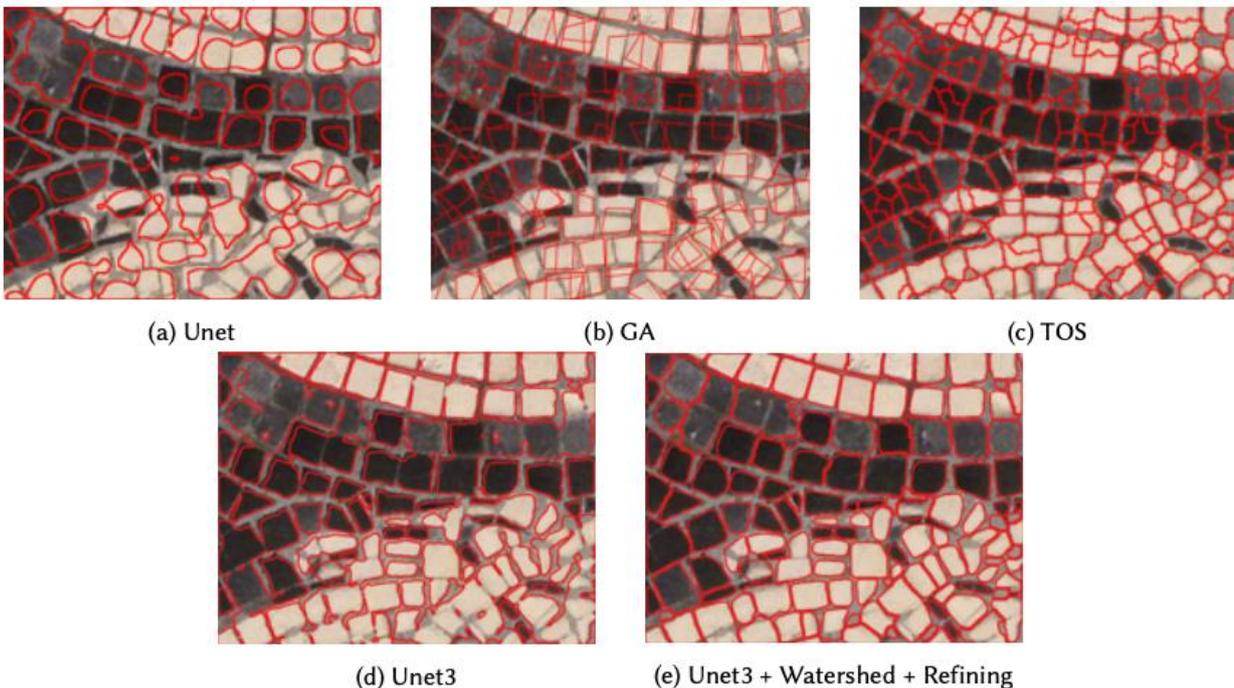
Figure 8: Tessera segmentation results.

segmentation (giving a qualitative ratio of the classes), without considering the number of objects. The metric we introduced, instead, is instance-based, hence it considers the numerosity of the objects (namely tesserae) and is computed as a ratio between the numerosity of the correctly predicted tesserae and the ground truth.

The correctly predicted tesserae are evaluated considering the ratio between overlapping areas and the area of the real tesserae. If this ratio overcomes the described thresholds (the overlap constraints), the tesserae are correctly predicted, otherwise discarded. However, this metric needs further investigation when used to evaluate perimeters' similarity. Given the above, it can be concluded that metrics proposed in (Fenu et al., 2015; Fenu et al., 2020) are less robust than the ones here proposed, as they consider both measures without fixing a pre-defined threshold.

Besides, readers could find useful a discussion about the execution time. Albeit the code has not optimized for enhancing its computational performances, from our estimation it can be deduced that Mo.Se. overcomes humans' limitations. Our algorithm, from neural network segmentation (model already trained) to refining takes about 15 minutes per square meter on a machine equipped with 2 Intel(R) Xeon(R) Silver 4214 CPU @ 2.20GHz and 128GB RAM. The average time of annotation (perimeter drawing only) of a fast and expert human operator is about 10 seconds per tesserae. If square meter of mosaic uniformly covered with average area tesserae is considered, a human operator would take more than 30 hours per square meter.

About its reliability, compared with a human operator, two different measures are considered: the capability to discriminate one tessera from another, the capability to detect the exact perimeter. In the first case, the accuracy of the human operator can be considered 100%. For this reason, the annotation of the human operator is considered ground truth. In the second case, the accuracy of the human operator depends on the time dedicated to the annotation, related to the number of points to approximate the polygon (vector), as well as the operator's perception due to the image definition. The algorithm has undoubtedly a lower accuracy than the human operator in discriminating one tessera from another and recognizing the perimeter of the tessera, but the annotation made by the human operator, for reasons of time that can be dedicated to the annotation, is approximate. For this reason, in many cases, the algorithm is more efficient.

Finally, an interesting argument of discussion is the image quality together with the complexity of the mosaic. Preliminary experiments were performed to evaluate how much the method is dependent on image resolution. The best results occurred, as expected, with the most defined image. The higher the image definition, the more textures and contrast between the edges are visible. Conversely, a less defined image reduces these features, making the task difficult even for the human operator. Tesserae's shape does not seem to affect results, whilst their size does. In fact, in most cases, when fragmented tesserae occur, individual fragments are classified as tesserae. In case tesserae are bigger instead, the algorithm can correctly classify tesserae. Colour greatly affects the accuracy of the method. This issue is not due to the colour itself, but rather to the shades that are emphasized with that colour. Shades are very visible in ruined tiles, and probably due to erosion or powder coating. For this reason, the algorithm can confuse them as a gap. Thus, this latter issue is strictly dependent on the orthoimage.

6. Conclusion and future works

Automatic segmentation of ancient mosaics can help archaeologists and CH experts to build digital collections and to automatically compare mosaics by database indexing and content-based retrieval tools. In this paper, Mo.Se. algorithm has been proposed, which comprises deep learning and image segmentation methods for the automatic segmentation of tesserae. The mosaic image segmentation pipeline gives segments corresponding to the single tessera. Experiments have been assessed on the pavement of the Byzantine mosaics of St. Stephen Church, in Jordan. Moreover, in order to validate and generalise Mo.Se., open access mosaic datasets have been used for the comparison. In particular, the following phases have been performed: HD acquisition of orthophoto RGB of mosaic pavement of St. Stephen's Church in Umm ar-Rasas in Jordan; Tesserae Segmentation using U-Net 3 Deep Neural Network; Tesserae Segmentation using U-Net 3 Deep Neural Network and Hierarchical Watershed Algorithm; Approach Performance comparison with SoA datasets. Our experimental analysis shows that Mo.Se. is tailored for the segmentation of ancient mosaics because outperforms SoA works for this task.

Mo.Se can be the core for the development of tools that can be of interest for museums, to digitalise mosaics for an inventory of theirs. The purpose is to facilitate the very time-consuming task of drawing the tesserae in CAD/GIS environment that, nowadays, is manually performed.

Future works will involve the definition of a procedure for the automatic recognition of background, figures and disfigurements (where occurs) in the scenes.

Other investigations will be devoted to the achievement of semantically enriched information and the extraction of objects with semantic meaning in a complex mosaic scene, thus providing CH experts with a tool for the automatic extraction of geometric and semantic information of ancient mosaics.

Acknowledgements

References

Bartoli, A., Fenu, G., Medvet, E., Pellegrino, F. A., & Timeus, N. (2016). Segmentation of mosaic images based on deformable models using genetic algorithms. In *International Conference on Smart Objects and Technologies for Social Good* (pp. 233–242). Springer, Cham. https://doi.org/10.1007/978-3-319-61949-1_25

Battiato, S., Di Blasi, G., Farinella, G. M., & Gallo, G. (2007). Digital mosaic frameworks - An overview. In *Computer Graphics Forum* (Vol. 26, No. 4, pp. 794–812). Oxford, UK: Blackwell Publishing Ltd. https://doi.org/10.1111/j.1467-8659.2007.01021.x

Beucher, S., & Lantuéjoul, C. (1979). Use of watersheds in contour detection. *International Workshop on Image Processing: Real-time edge and motion detection/estimation.* Rennes, France.

Benyoussef, L., & Derrode, S. (2011). Analysis of ancient mosaic images for dedicated applications. *Digital Imaging for Cultural Heritage Preservation: Analysis, Restoration, and Reconstruction of Ancient Artworks*, 385.

Bonfigli, R., Felicetti, A., Principi, E., Fagiani, M., Squartini, S., & Piazza, F. (2018). Denoising autoencoders for non-intrusive load monitoring: improvements and comparative evaluation. *Energy and Buildings*, *158*. https://doi.org/10.1016/j.enbuild.2017.11.054

Bordoni, L., & Mele, F. (Eds.). (2016). *Artificial intelligence for cultural heritage*. Cambridge Scholars Publishing.

Bourke, P. (2014). Novel imaging of heritage objects and sites. In *2014 International Conference on Virtual Systems & Multimedia (VSMM)* (pp. 25–30). IEEE. https://doi.org/10.1109/VSMM.2014.7136666

Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T., & Ronneberger, O. (2016). 3D U-Net: learning dense volumetric segmentation from sparse annotation. In *International conference on medical image computing and computer-assisted intervention* (pp. 424-432). Springer, Cham. https://doi.org/10.1007/978-3-319-46723-8_49

Cipriani, L., & Fantini, F. (2017). Digitalization culture vs archaeological visualization: integration of pipelines and open issues. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, *42*, 195.

Djibril, M. O., & Thami, R. O. H. (2008). Islamic geometrical patterns indexing and classification using discrete symmetry groups. *Journal on Computing and Cultural Heritage (JOCCH)*, *1*(2), 1–14. https://doi.org/10.1145/1434763.1434767

Djibril, M. O., Thami, R. O. H., Benslimane, R., & Daoudi, M. (2005). Une nouvelle technique pour l'indexation des arabesques basée sur la dimension fractale. *Univ. Mohamed V, Maroc*.

Falk, T., Mai, D., Bensch, R., Çiçek, Ö., Abdulkadir, A., Marrakchi, Y., Böhm, A., Deubner, J., Jäckel, Z., Seiwald, K., & Dovzhenko, A. (2019). U-Net: deep learning for cell counting, detection, and morphometry. *Nature methods*, *16*(1), 67–70. https://doi.org/10.1038/s41592-018-0261-2

Felicetti, A., Albiero, A., Gabrielli, R., Pierdicca, R., Paolanti, M., Zingaretti, P., & Malinverni, E. S. (2018). Automatic Mosaic Digitalization: a Deep Learning approach to tessera segmentation. In *METROARCHEO, IEEE International Conference on Metrology for Archaeology and Cultural Heritage. Cassino*. https://doi.org/10.1109/MetroArchaeo43810.2018.13606

Fenu, G., Jain, N., Medvet, E., Pellegrino, F. A., & Namer, M. P. (2015). On the assessment of segmentation methods for images of mosaics. In *VISAPP (3)* (pp. 130–137). https://doi.org/10.13140/RG.2.1.3025.6489

Fenu, G., Medvet, E., Panfilo, D., & Pellegrino, F. A. (2020). Mosaic images segmentation using U-net. In *International Conference on Pattern Recognition Applications and Methods* (pp. 485–492). Scitepress. http://dx.doi.org/10.5220/0008967404850492

Fontanella, F., Molinara, M., Gallozzi, A., Cigola, M., Senatore, L. J., Florio, R., Clini, P., & Celis, F. (2019). HeritageGO (HeGO). A social media based project for cultural heritage valorization. In *Adjunct Publication of the 27th Conference on User Modeling, Adaptation and Personalization* (pp. 377–382). https://doi.org/10.1145/3314183.3323863

Gil, F. A., Gomis, J. M., & Pérez, M. (2009). Reconstruction techniques for image analysis of ancient islamic mosaics. *International Journal of Virtual Reality*, *8*(3), 5–12. https://doi.org/10.20870/IJVR.2009.8.3.2735

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Kohl, S., Romera-Paredes, B., Meyer, C., De Fauw, J., Ledsam, J. R., Maier-Hein, K., Eslami, S.M.A, Rezende, D.J., & Ronneberger, O. (2018). A probabilistic U-Net for segmentation of ambiguous images. In *Advances in Neural Information Processing Systems* (pp. 6965–6975). https://arxiv.org/abs/1806.05034

Liciotti, D., Paolanti, M., Pietrini, R., Frontoni, E., & Zingaretti, P. (2018). Convolutional networks for semantic heads segmentation using top-view depth data in crowded environment. In *24th international conference on pattern recognition (ICPR)* IEEE. https://doi.org/10.1109/ICPR.2018.8545397

Maghrebi, W., Ammar, A. B., Alimi, A. M., & Khabou, M. A. (2013). An intelligent multi-object retrieval system for historical mosaics. *Editorial Preface*, *4*(4). https://doi.org/10.14569/IJACSA.2013.040417

Maghrebi, W., Baccour, L., Khabou, M. A., & Alimi, A. M. (2007). An indexing and retrieval system of historic art images based on fuzzy shape similarity. In *Mexican International Conference on Artificial Intelligence* (pp. 623–633). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-76631-5_59

Maghrebi, W., Borchani, A., Khabou, M. A., & Alimi, A. M. (2007). A system for historic document image indexing and retrieval based on XML database conforming to MPEG7 standard. In *International Workshop on Graphics Recognition* (pp. 114–125). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-88188-9_12

Malinverni, E. S., Pierdicca, R., Di Stefano, F., Gabrielli, R., & Albiero, A. (2019). Virtual museum enriched by GIS data to share science and culture. Church of Saint Stephen in Umm Ar-Rasas (Jordan). *Virtual Archaeology Review*, *10*(21). https://doi.org/10.4995/var.2019.11919

M'hedhbi, M., Mezhoud, R., M'hiri, S., & Ghorbel, F. (2006). A new content-based image indexing and retrieval system of mosaic images. In *2006 2nd International Conference on Information & Communication Technologies* (Vol. 1, pp. 1715–1719). IEEE. https://doi.org/10.1109/ICTTA.2006.1684644

Pierdicca, R., Frontoni, E., Malinverni, E. S., Colosi, F., & Orazi, R. (2016). Virtual reconstruction of archaeological heritage using a combination of photogrammetric techniques: Huaca Arco Iris, Chan Chan, Peru. *Digital Applications in Archaeology and Cultural Heritage*, *3*(3). https://doi.org/10.1016/j.daach.2016.06.002

Pierdicca, R., Frontoni, E., Zingaretti, P., Malinverni, E. S., Colosi, F., & Orazi, R. (2015, August). Making visible the invisible. Augmented reality visualization for 3D reconstructions of archaeological sites. In *International Conference on Augmented and Virtual Reality*. Springer, Cham. https://doi.org/10.1007/978-3-319-22888-4_3

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention* (pp. 234–241). Springer, Cham. https://doi.org/10.1007/978-3-319-24574-4_28

Vincent, L., & Soille, P. (1991). Watersheds in digital spaces: an efficient algorithm based on immersion simulations. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (6), 583–598. https://doi.org/10.1109/34.87344

Youssef, L. B., & Derrode, S. (2008). Tessella-oriented segmentation and guidelines estimation of ancient mosaic images. *Journal of Electronic Imaging*, *17*(4), 043014. https://doi.org/10.1117/1.3013543

Zarghili, A., Gadi, N., Benslimane, R., & Bouatouch, K. (2001). Arabo-Moresque decor image retrieval system based on mosaic representations. *Journal of Cultural Heritage*, *2*(2), 149–154. https://doi.org/10.1016/S1296-2074(01)01116-5

Zarghili, A., Kharroubi, J., & Benslimane, R. (2008). Arabo-Moresque decor images retrieval system based on spatial relationships indexing. *Journal of cultural heritage*, *9*(3), 317–325. https://doi.org/10.1016/j.culher.2007.10.008

Zitová, B., Flusser, J., & Šroubek, F. (2004). An application of image processing in the medieval mosaic conservation. *Pattern analysis and applications*, *7*(1), 18–25. https://doi.org/10.1007/s10044-003-0200-3