Automatic detection of stone pavement's pattern based on UAV photogrammetry

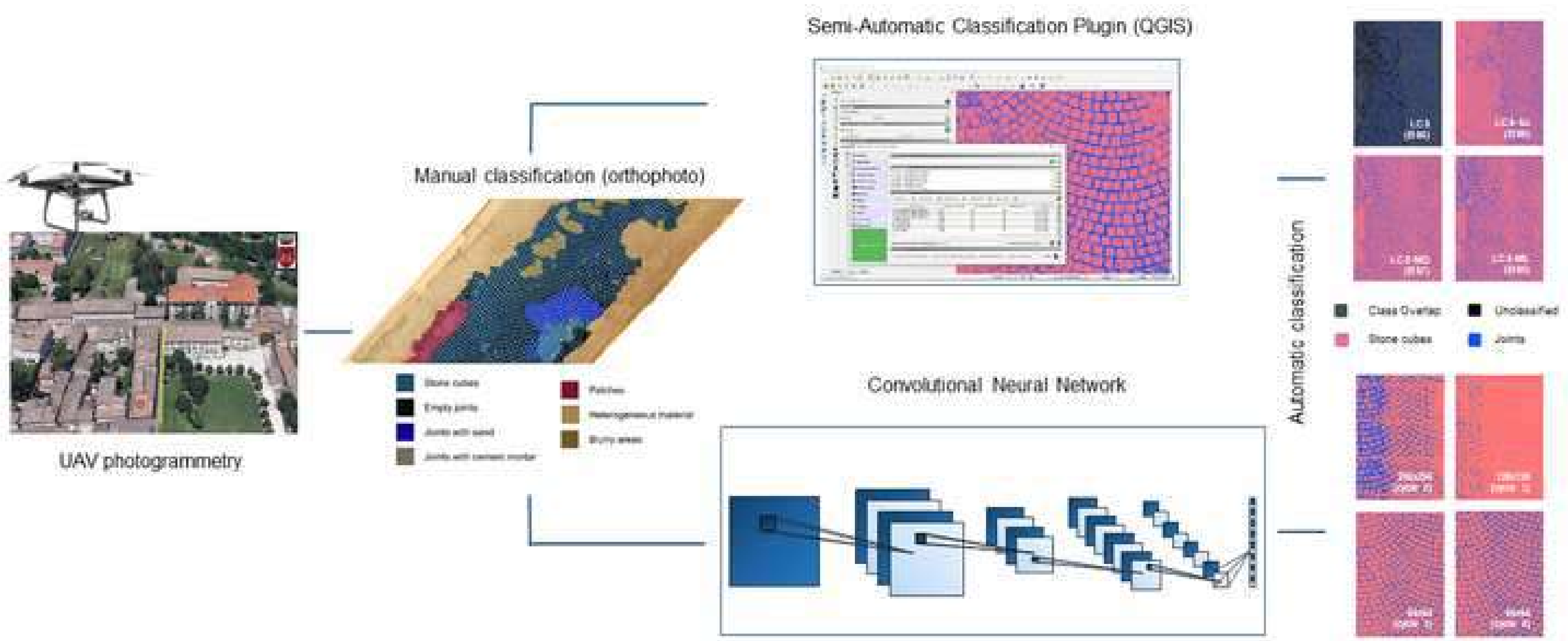
Original
Automatic detection of stone pavement's pattern based on UAV photogrammetry / Garilli, E.; Bruno, N.; Autelitano, F.; Roncella, R.; Giuliani, F.. - In: AUTOMATION IN CONSTRUCTION. - ISSN 0926-5805. - 122:103477(2021), pp. 1-14. [10.1016/j.autcon.2020.103477]

Availability:
This version is available at: 11381/2886020 since: 2021-09-23T19:28:17Z

Publisher:
Elsevier B.V.

Published
DOI:10.1016/j.autcon.2020.103477

Graphical Abstract



Automatic detection of stone pavement's pattern based on UAV photogrammetry

Erika Garilli[a], Nazarena Bruno[a], Federico Autelitano[a], Riccardo Roncella[a], Felice Giuliani[a]

[a]Dipartimento di Ingegneria e Architettura, Università di Parma, Parco Area delle Scienze 181/A Parma 43124, Italy

Corresponding author:

Federico Autelitano, Dipartimento di Ingegneria e Architettura, Università di Parma, Parco Area delle Scienze 181/A Parma 43124, Italy. Tel. +390521905971. E-mail address: federico.autelitano@unipr.it

Abstract

Pavement management system (PMS) is a set of tools that assist road agencies in finding optimal strategies for maintaining pavements in a serviciable condition over a period of time. Usually, municipalities base their PMS on the deterioration monitoring through a visual survey but the distresses identification is complex and the operations are based on visual and instrumental inspections. As regards natural stone pavements, which are very widespread in the road heritage of cities, in literature there are very few studies. The authors analyzed two supervised classification approaches (Semi-Automatic Classification Plugin for QGIS and a Convolutional Neural Network (CNN)), based on Unmanned Aerial Vehicle (UAV) photogrammetry, to detect stone pavement's pattern. This study showed that using a U-Net CNN on images obtained from UAV is an excellent alternative to the traditional manual inspection and can be implemented for other types of stone pavements, also with the aim of distress identification.

Highlights

- PMS of urban road rarely has an organized structure.
- There is a lack of tools for natural stone pavements distress analysis.
- Two supervised classification approaches based on UAV photogrammetry were analyzed.
- Supervised classifications are used for the detection of stone pavement's pattern.
- U-Net CNN is an excellent alternative to the manual inspection.

Keywords

Pavement management system, stone pavement, segmental pavement, automatic classification, deep learning, convolutional neural network

1. Introduction

Pavement management system (PMS) is a set of tools that assist road agencies in finding optimal strategies for maintaining pavements in a serviciable condition over a period of time. The implementation of PMS is a complex operation with the aim of analyzing and modelling road surface deterioration, providing appropriate road maintenance and rehabilitation strategies, maximizing performance during the service life establishing priority scheduling optimizing agency costs [1–3]. It is highly data dependent and acquiring these data is expensive and time-consuming as it has seen the involvement of increasingly specialised skills and equipment over time. The PMSs are differently graduated on the basis of the extension and type of road-network to be managed, as well as on the budget and instrumental investigation and processing systems available. However, most of the PMSs were created to manage large network referring to major road and airport infrastructures while the applications of such systems to urban areas are few in the world [4–8]. The PMS of urban road rarely has, for several reasons, an organized structure. Primarily, they are subject to economic constraints, but also, because of the huge diversification of type and use of roads, they are subject to different traffic conditions with too short extensions for systematic long-term investment planning. Usually, municipalities base their PMS on the monitoring of the deterioration through a visual survey such as the pavement condition index (PCI) assessment procedure based on a numerical scale from 100 (perfect condition) to 0 (failed pavement) [9]. The use of the PCI in urban areas has considerable limits because the distress manual identification is more complex and the operations are basically based on visual and instrumental inspections on short paving trunks. The distresses are identified from specific typological catalogues, in a not easy environment, with many obstacles and operational problems typical of urban life (parking, facilities, etc..).

With the exception of a few isolated studies [10], the literature on PMS does not cover situations that are very widespread in the road heritage of cities, i.e. natural stone pavements. They represent the largest framework of the historical European city's material structure, the main features of the architecture of the public roads, right up to the development of motor vehicles [11,12]. Stone block pavements, falling within the segmental pavements category, differ from others road pavements because the wearing course is made of individual small elements placed in a predefined laying pattern, above an unbound or bound bedding layer. Stone elements characterized by different materials, shapes and dimensions can be used for the surface layer. Under this course the structure is similar to that of a flexible or semi-rigid pavement. The structural capacity of such pavements, subject to both vertical and horizontal loads due to braking, acceleration and steering, depends not only on underlying layers but also on the size of the elements used, on the laying patterns, on the joints' filling material and on joints' thickness [13]. The main types of deterioration of segmental pavements can be divided into two macro-categories [14]: vertical displacements (depressions, faulting, heave and rutting), which are mainly due to the high vertical loads on the pavement together with a lack of bearing capacity of the deeper layers, and horizontal displacements due to the horizontal component of vehicular loads transmitted under conditions of adherence to the pavement which causes the relocation of the blocks (horizontal creep) [15]. A third type of distress concerns the fracture of the blocks due to the incorrect selection of used materials. However, an interaction between the causes of deterioration is recurrent with the generation of other types of distress such as excessive joints width, joint filling loss and pumping, missing pavers and patching [10,14,16]. Stone paving involves a multiplicity of aspects to be evaluated that are not only related to the maintenance of structural and functional requirements, but also aesthetic and formal: poorly executed repairs, joint emptying and loss of laying pattern geometry are themselves important distresses hard to detect and quantify with acceptable speed and precision. The traditional manual inspection of pavements, especially for segmental pavements, can be a complex task for a person and the supervision usually requires trained staff, becoming highly time-consuming, labor-intensive, subjective and quite expensive.

Thus, the possibility of using different methodologies for the automatic detection of a segmental pavement's pattern, which is the main objective of this study, would represent an unexplored and interesting strategy to be used into an urban PMS. Recent studies are attempting to automate the process of analysis of road pavements and the detection of distresses in order to obtain accurate and low-cost data, but they mainly refer to asphalt and concrete pavements, which are the most common and congested roads where distresses can significantly affect road users' safety and comfort [17]. The most studied techniques involve laser-based systems and imagery from cameras. The formers include a variety of devices such as laser profilers and terrestrial laser scanners but are generally very expensive. For this reason, systems based only on imagery are considered a viable alternative at a significantly lower cost. After acquisition, images are processed to analyze distresses. As far as data post-processing is concerned, traditional 2D image post-processing techniques are, for instance, edge detection and morphology, binarization and thresholding [18]. Applications for automated or semi-automatic data extraction are instead based on image processing algorithms and computer vision [19–22] or emerging methods such as deep learning [23–26]. Deep learning and semi-automatic classification tools, has shown good results in automatically detecting and assessing the health condition of civil infrastructure such as flexible pavements [27–30]. In the European and North American contexts, however, in many urban centers there is a strong presence of streets with historical pavements, such as stone cubes pavements, that require attention. For this type of pavements, in fact, even before the survey of distresses, it is fundamental the laying pattern detection. Although laying pattern is one of the aspects that more than others characterize stone pavements but, at the same time, it is difficult to evaluate its geometry as well as its geometric stability over time.

The attempt to automatically detect a historical natural stone pavement in this study has included a preliminary survey conducted using photogrammetry from Unmanned Aerial Vehicle (UAV) and a subsequent post-processing phase. UAV technique is not commonly used for road pavement surveying, but it is now well established in other civil engineering applications. As will be highlighted in this paper, the use of UAVs allows the rapid acquisition of images compared to a ground survey, high resolution images thanks to the sensors used, which are increasingly high performance, the integrated use of radiometric and geometric 3D information, the possibility of immediate georeferencing (and therefore the location of any distress) thanks to GPS sensors available. For the automatic pattern detection, two supervised classification approaches will be tested. The former is based on the use of algorithms for the classification of remote sensing images. In particular the Semi-Automatic Classification Plugin (SCP) [31], a free open source plugin for QGIS [32], was used. The choice of this tool is due to the widespread use of QGIS software by municipalities. The latter method instead relies on the use of a Convolutional Neural Network (CNN), which allows a higher degree of automation and to overcome some limitations related to the radiometric uncertainties present in the RGB input data to be analyzed. The combined use of UAV photogrammetry and semi-automatic classifications can increase the efficiency of the process by outperforming people in speed and accuracy, working evenly and being independent of human factors.

2. Materials and methods

2.1. Study area

The survey was carried out on a natural stone pavement located in the historical city center of Ascoli Piceno (Marche region of Italy). The surveyed street is “Via Pietro Alamanni” (the yellow line in Figure 1), that is parallel to the *cardo maximus* (Via Cassero, Via Malta and Via Pretoriana, the green line) and intercepts at right angles “Corso Mazzini” (the red line) which represents the *decumanus maximus* of the city, as Ascoli Piceno is a city of Roman structure. The analyzed street is within a Limited Traffic Zone that allows transit only to residents and is also open to unauthorized vehicles in exceptional periods of time. The street geometry allows the transit of small vehicles, so that only cars and two-wheeled vehicles are allowed to pass through; therefore, it can be assumed that there are low transits and loads not greater than those of cars.

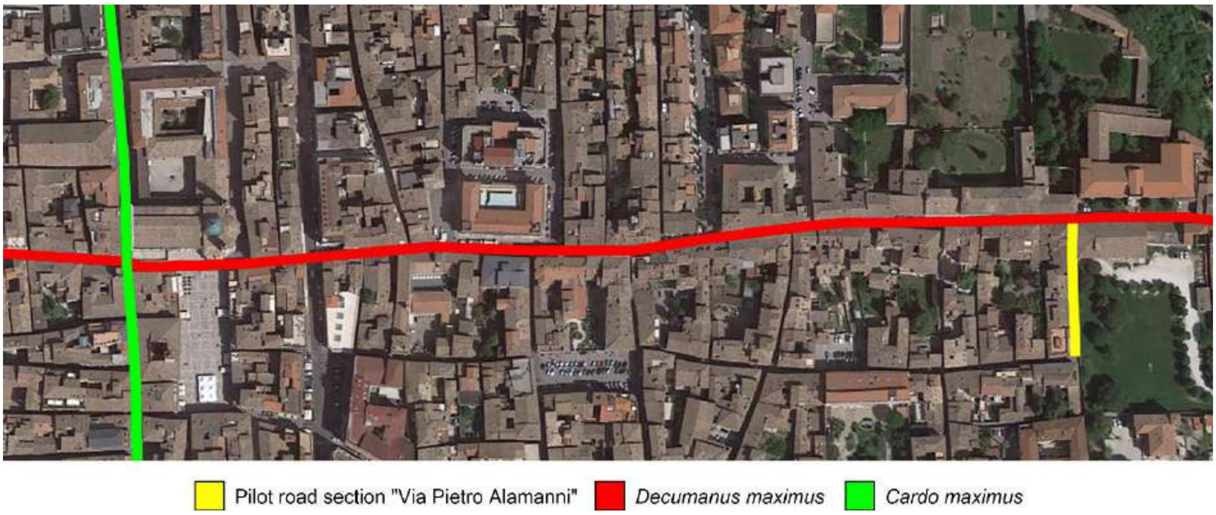


Figure 1 – Historic center of Ascoli Piceno and identification of the pilot road section.

The historical center of the city of Ascoli Piceno is strongly characterized by the presence of stone pavements of different types; the street surveyed specifically consists of small, cubic, Trentino’s porphyric elements, commercially in 6/8 class placed in an overlapping arcs laying pattern (Figure 2). Table 1 summarizes the property of the commercial 6/8 class. The choice of this street as the object of the survey derives from two main aspects. The first one is linked to the fact that the historic center of Ascoli Piceno, almost entirely built in travertine, is among the most admired in Marche region and central Italy, due to its artistic and architectural heritage and has been nominated several times for the list of UNESCO World Cultural and Natural Heritage Sites. Secondly, “Via Pietro Alamanni” had a high level of deterioration, so it is of great interest in the issue of stone pavement distresses. Moreover, in view of its future reconstruction, it has been possible to use it as a pilot road section on which various types of distresses survey have been carried out as well as a set of destructive and non-destructive tests.

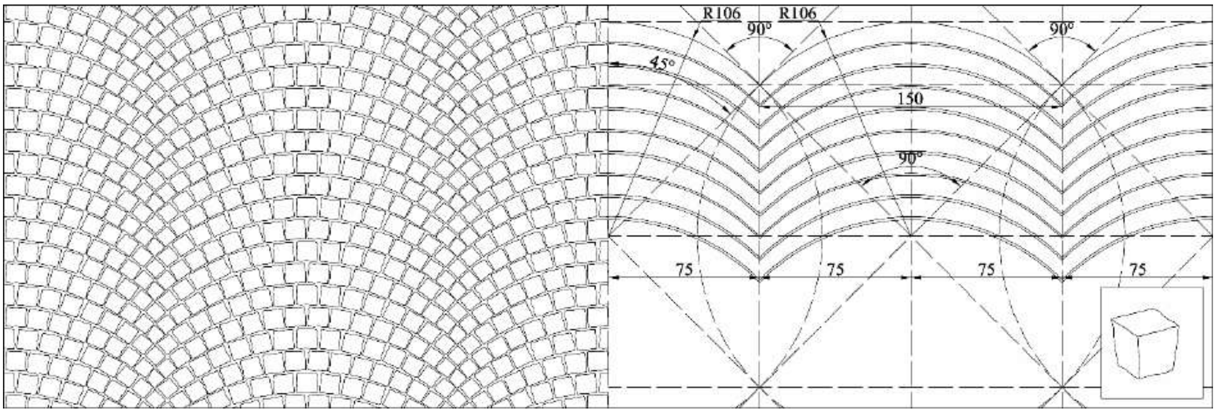


Figure 2 - Small, cubic, Trentino’s porphyric elements in 6/8 class placed in an overlapping arcs laying pattern (dimensions in cm).

Table 1 - Property of the commercial 6/8 class.

Length and width	6.0-9.0 cm
Height	5.5-8.0 cm
Weight	130-135 kg/m^2
Number of elements	approx. 155-160/m^2

As shown in Figure 3, the pavement presented several distresses. The pavement had numerous patches: Figure 3a displays sections of pavement where there are missing pavers which have been reinstated with a dissimilar material. With regard to the joints, i.e. the empty spaces between adjacent stone cubes, they were filled with different materials: as can be seen in Figure 3b some of these still had the original filling material, i.e. sand, others were filled with cement mortar, others have lost the filling material and have not been restored. In addition, there were areas in which the joints between blocks have widened. Excessive joint width can occur from a number of factors and as joints get wider, the block layer becomes less stiff and can lead to overstressing the substructure layers or to the loss of some stone cubes that have not been replaced and have not been patched. Finally, in many areas (Figure 3a) the stone pavement showed numerous vertical irregularities like depressions (sections that present lower elevations than the surrounding areas), faulting (areas where the elevation of adjacent stone cubes differ or have rotated), heave (sections that have elevations that are higher than the surrounding areas) and rutting (a surface depression in the wheel path).

a) b)

Figure 3 – Distresses detected on the stone pavement during the survey.

2.2. Data acquisition and photogrammetric processing

The survey was carried out on a pilot road section, with a length of about 80 m, by drone photogrammetry and involved a surface equal to 472 m^2. All street images were collected by a DJI Phantom 4 pro quadcopter drone. Thirteen control points, detected by Topcon ISO1 Total Robotic Station and GPS antennas, were used for the definition of the reference frame and for the optimization of the image orientation solution. The targets were placed along the sides of the street at an average distance of 8 meters to ensure a constant distribution along the entire planimetric extension of the street. A total of 194 images were acquired, alternating nadir and oblique shots. The equipped camera has one-inch 20-MP (5472 × 3648 resolution) CMOS sensor with a focal length of 8.8 mm. The shutter speed is 1/2000 to 1/8000 s and the sensor size is 12.83 × 7.22 mm. The actual ground resolution of the acquired images can be quantified using the Ground Sampling Distance (GSD, i.e. the size of an object element corresponding to a single pixel in the digital image). The GSD can be calculated as in Eq.(1) [33]:

$$GSD = \frac{Z \cdot p}{f} \qquad (1)$$

where Z is the object distance (distance from the camera to the pavement surface), p is the pixel size of the sensor and f is the focal length of the lens. In this study GSD was equal to 1.3 mm.

The expected depth (along Z direction) accuracy σ can be estimated by the Eq. (2):

$$\sigma = \frac{Z}{f} \cdot \frac{Z}{B} \cdot \sigma_m \qquad (2)$$

where σ_m is the measurement precision of the image coordinates (assumed to be ±1 pixel) and B is the base length (distance between the two consecutive shots) [34]. The resulting depth precision of coordinates is about 3.1 mm.

The first processing stage (3D reconstruction and orthophoto generation) was carried out with Agisoft Metashape software; the software adopts a fairly standardized processing pipeline: image block orientation through a structure from motion automatic procedure, generation of a 3D point cloud representing the detected object, generation of a triangular mesh model from the point cloud, creation of raster products such as Digital Elevation Model (DEM) and orthophotos. Considering the GSD, the orthophotos have been generated with 2 mm per pixel resolution, suitable to better appreciate the texture of the stone pavement. To achieve the highest possible resolution, all processing steps have been done using the highest quality settings offered by the software. Two sections of the orthophoto obtained from the survey carried out on “Via Pietro Alamanni” are shown in Figure 4.

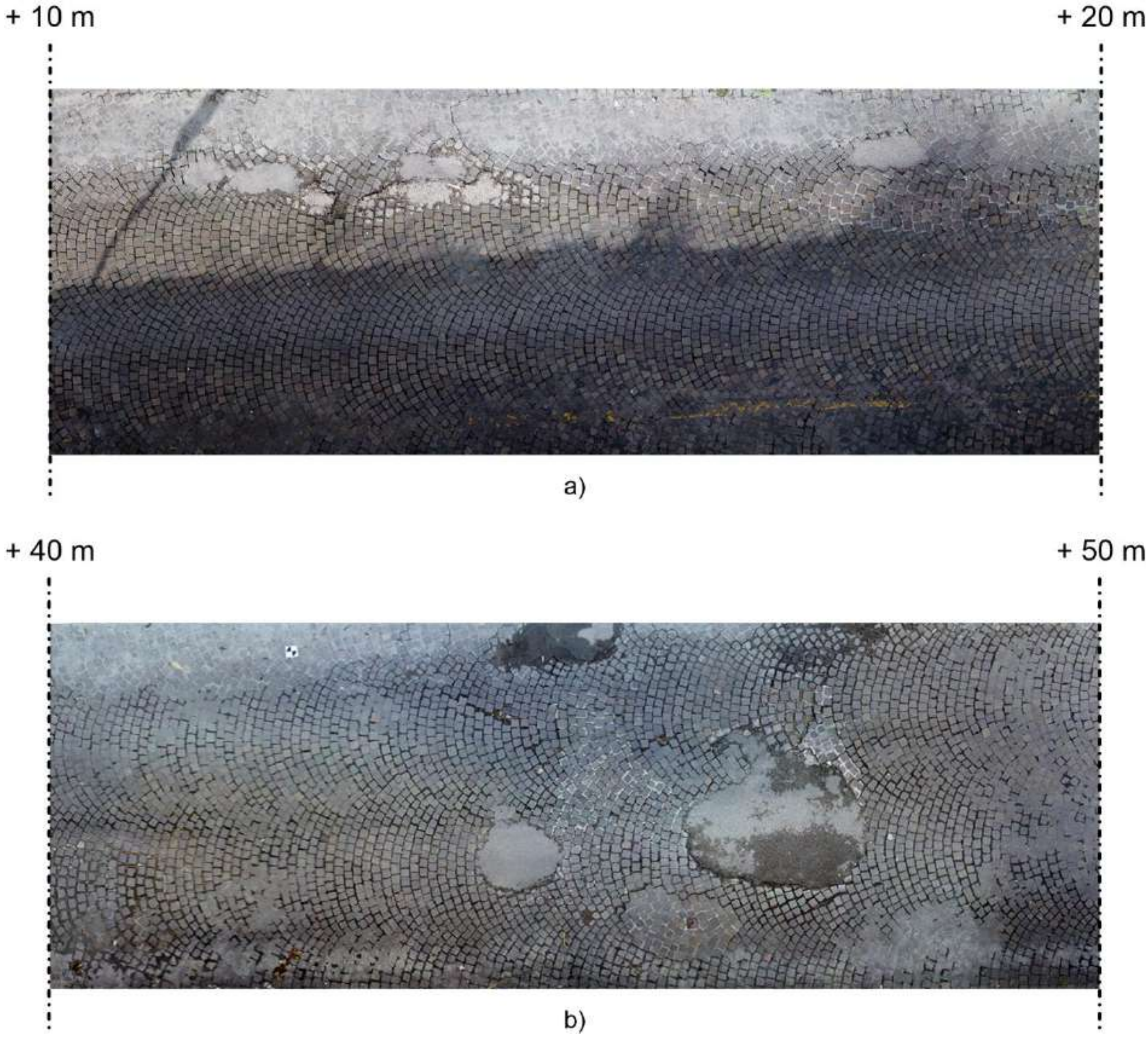


Figure 4 – Sections of the orthophoto obtained from the survey carried out in “Via Pietro Alamanni”; a) In the orthophoto the strong presence of shadows is evident; b) In this section patches with different material filling are present.

2.3. Training area and class identification

Both classification methodologies considered in this work (SCP and CNN) are supervised classification techniques. As such, they consist of two successive phases: the former (*training*) involves the classifier training on the basis of data provided by the operator; the latter (*classification*) consists of classifying the entire dataset on the basis of the initial training. The training phase is computed on a controlled area, starting from data manually classified by the user. In this case, the manual classification for training was made on a portion of the entire orthophoto (among 25 m^2). The selected area had homogeneous lighting, was well representative of the characteristics of the pavement and contained a significant sample of the laying pattern (with stone cubes and joints), as well as of the materials (stone cubes, sand, cement mortar) and distresses (mainly patches and missing cubes). The manual classification was made in QGIS by vector drawing of polygons representing the different materials on stone pavements. Macroclasses and the related classes have been identified and shown in Table 2 and Figure 5.

Table 2 – Macroclasses and classes

Macroclass	Class
Pavement	Stone cubes
	Patches
Joints	Empty joints
	Joints with sand
	Joints with cement mortar
Other	Heterogeneous material
	Blurry areas

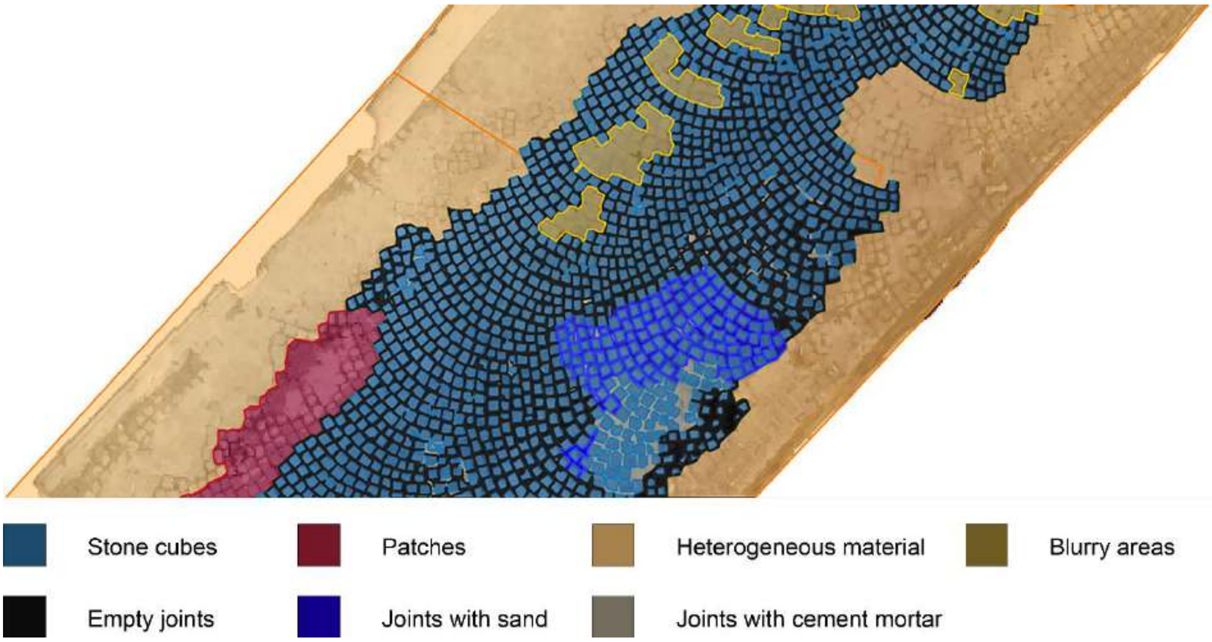
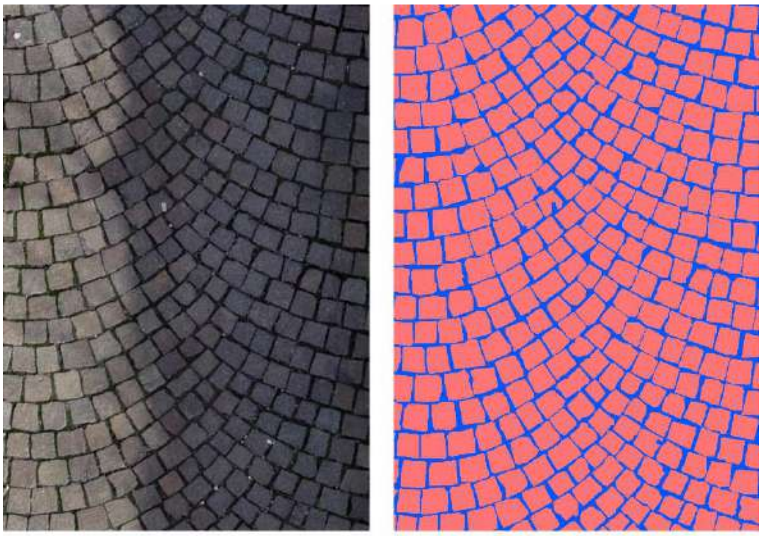


Figure 5 – Manual classification of a portion of orthophoto

However, since the purpose of this study was the automatic identification of the laying pattern and thus the distinction between stone cubes and joints, only the two classes of stone cubes and joints were considered, without taking into account for the latter the material differences and looking only at their geometric characteristics. The manually classified dataset used as training input was the same for the two tested methodologies (SCP and CNN).

2.4.Testing area and data evaluation

After the classification phase, the output results were validated by comparison with a reference dataset in order to validate them and assess their accuracy. The validation was done on another portion of orthophoto (area 3 m^2 circa), a sample of which is represented in Figure 6. Also in this area a manual classification was done according to the same methodology described in the paragraph 2.3, but identifying only stone cubes and joints without other elements such as patches, in order to have a reference dataset. The manual classification, in fact, is the most onerous methodology, but at the same time it is the most accurate and the most reliable reference for evaluating the accuracy of automatic algorithms. So, the objective of these tests is to obtain, with automatic algorithms, the closest results to the one provided by manual classification.



a) b)

Figure 6 - Training validation input: a) image to classify and b) reference raster image resulting from manual classification.

For accuracy evaluation, reference was made to traditional metrics with four possible types of outcomes concerning the assessments of street segments given by the classification system: true positive (TP - the analyzed pixel belongs to stone cubes and the system correctly detects it), true negative (TN – the analyzed pixel does not belong to stone cubes and the system correctly detects it), false positive (FP – the analyzed pixel does not belong to stone cubes but the system classifies it as a stone cube) and false negative (FN – the analyzed pixel belongs to stone cubes but the system does not classify it as a stone cube). Based on these outcomes, the recall, selectivity, precision, accuracy, F-score and Matthews correlation coefficient (MCC) metrics have been calculated to evaluate the classification results. The recall measures the proportion of actual positives that are correctly identified as such (Eq. (3)), while the selectivity measures the proportion of actual negatives that are correctly identified as such (Eq. (4)).

$$Recall = \frac{TP}{TP+FN} \tag{3}$$

$$\text{Selectivity} = \frac{TN}{TN+FP} \tag{4}$$

The precision (Eq. (5)) is the proportion of predicted positive that are true positive:

$$Precision = \frac{TP}{TP+FP} \tag{5}$$

The accuracy describes the percentage of the test data that are correctly classified (Eq. (6)):

$$Accuracy = \frac{number\ of\ correct\ predictions}{total\ number\ of\ predictions} \tag{6}$$

In this case we have a binary classification and accuracy can also be calculated in terms of positives and negatives as in Eq. (7):

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{7}$$

Accuracy works well only if there are equal number of samples belonging to each class; in this case the classes are unbalanced, and it is therefore necessary to analyze other parameters like F-score and Matthews correlation coefficient (MCC).

The F_1 score is the harmonic mean of the precision and recall (Eq.(8)):

$$F_1 score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \tag{8}$$

Finally, the MCC (Eq. (9)) is the most significant coefficient in a binary classification in which the classes are of very different size:

$$MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP+FP)\cdot(TP+FN)\cdot(TN+FP)\cdot(TN+FN)}} \tag{9}$$

This coefficient considers true and false positives and negatives and is generally regarded as a balanced measure which can be used even if the classes are of very different sizes. The MCC is a correlation coefficient between the observed and predicted binary classifications; it returns a value between −1 and +1. A coefficient of +1 represents a perfect prediction, 0 no better than random prediction and −1 indicates total disagreement between prediction and observation.

2.5. Supervised automatic classification

2.5.1. Semi-Automatic Classification Plugin

The Semi-Automatic Classification Plugin (SCP) is a free open source plugin for QGIS that allows for the supervised classification (or semi-automatic classification) of remote sensing images. It provides several tools for raster processing (images download, preprocessing, post processing, raster calculation) to perform the land cover classification. A supervised classification is a machine learning technique that, in this case, allows the identification of materials in an image based on their spectral signatures. Each material has a unique signature, i.e. the reflectance as a function of wavelength [35] which can be used for material classification. To start the process, the user is required to select one or more training areas (or regions of interest, ROI) for each land cover class identified in the image. ROIs are polygons drawn on homogeneous areas of the image that include pixels belonging to the same class. The classifier then compares the spectral signatures of the train elements with those of the elements in the image to be classified. In this case study, the ROIs were identified during the previously described manual classification and the spectral signature of reference land cover classes are calculated considering the values of pixels inside each ROI belonging to the same class. The orthophoto to classify is an image in the visible field, i.e. for each pixel only the values concerning the RGB (Red, Green, Blue) bands are given. Since spectral signatures depend on the radiometric characteristics of the image, lighting conditions strongly influence spectral signatures identification, leading to the estimation of different signatures for the same material depending on shadow or sun exposure. Moreover, using only the three visible bands may lead to the misclassification of materials which are different but with similar radiometric characteristics in the visible field. To cope with this issues, emerging UAV systems allow to acquire also the near infrared component (NIR) (for instance the Anafi drone by Parrot), which could lead to improved results. However, without the availability of NIR data in these tests and to make the classification more robust against changing lighting condition in the scene (i.e. shadows or direct sun-light exposed areas), different pre-processing strategies have been adopted to maximize the differences between the ground cover classes. As visible in Figure 4a, which shows two different sections of the road to classify, the orthophotos have different lighting characteristics along the road extension, with the presence of sunny and shaded areas. The training area was identified on a portion of orthophoto with diffuse illumination and without the presence of sharp shadows (Figure 7a) to reduce the effects of different lighting conditions on the spectral signature calculation. With regard to the test areas (Figure 7b), instead, a preprocessing stage was made to mitigate shadows effect and equalize the global luminosity of the image with the reference image used for training. The shadow correction algorithms available in the literature are many [36–39]. In this study a locally adaptive filter was used, which ensure that, within a specified sliding window, the mean values over the three RGB channels match the global mean value of the reference image. It reduces uneven lighting, increasing RGB values of darker parts and decreasing RGB values of brighter areas (Figure 7c).

a) b) c)

Figure 7 – Image a) A portion of the training area used as reference image for equalization. Image b) A portion of test area where shadows are evident. Image c) The same portion of test area as image b after equalization process.

The algorithm performance is strongly influenced by the sliding window size used. The window size determines algorithm sensibility and, consequently, its capability of removing even small shadows (such as people's silhouettes, poles, fences etc.). On the other hand (as shown in Figure 8), using small sized window increases algorithm sensibility but decreases the resulting image contrast. So, with such algorithms, the correct balancing between precision in shadow detection and image contrast have to be strongly considered, not to compromise the classification stage. A sliding window of 50x50 pixels was used in this tests, since it ensured the best balancing between shadow removal and image contrast.

a) b)

Figure 8 – Comparison between equalized images obtained using different window sizes. Image a) Image is equalized using a window of 30x30 pix; Image b) Image is equalized using a window of 100x100 pix.

In addition, the conversion of the RGB images into HSV (Hue, Saturation and Value) space was performed. HSV [40] is a transformation of the RGB color space that gives a better separation of chromaticity and intensity. Hue is basically the color expressed from red to magenta as a number from 0 to 360 degrees, saturation describes the amount of grey in a particular color, value works in conjunction with saturation and describes the brightness or intensity of the color. Shadows mainly affect the value component, while for the same material the hue value should be basically the same regardless the shadows [41]. Thus, the conversion in HSV space was applied both in combination and as an alternative to shadow correction by image equalization, in order to improve the results provided by the equalization algorithm (especially in areas with still remaining shadows) and to verify whether the independent analysis of hue and value components can exclude the shadows effect from the spectral signature calculation.

In addition to lighting issues, the analyzed joints and stone cubes are composed by materials with similar radiometric characteristics, which makes it difficult to identify them unambiguously. The spectral signatures identified by considering only the three RGB bands had large overlapping portions and therefore would not have ensured a correct identification of the laying pattern. The RGB orthophotos were pre-processed by performing a Principal Component Analysis (PCA) to cope with this issue. PCA is a method for reducing the dimensions of measured variables (in this case bands) to the principal components, providing a new set of bands which are uncorrelated. This involves a linear transformation of the variables that projects the original ones into a new Cartesian system in which the new variable with the greatest variance is projected on the first axis, the new variable, second for variance value, on the second axis and so on. The reduction of complexity is limited to analyzing the main variables, by variance, among the new variables [42]. The variable with the highest variance is therefore represented by PCA band 1 and, as can be seen from the Figure 9 it increases the color differences between stone cubes and joints, so that joints appear almost white (high grey scale values) and the stone cubes almost black (low values), therefore more recognizable.

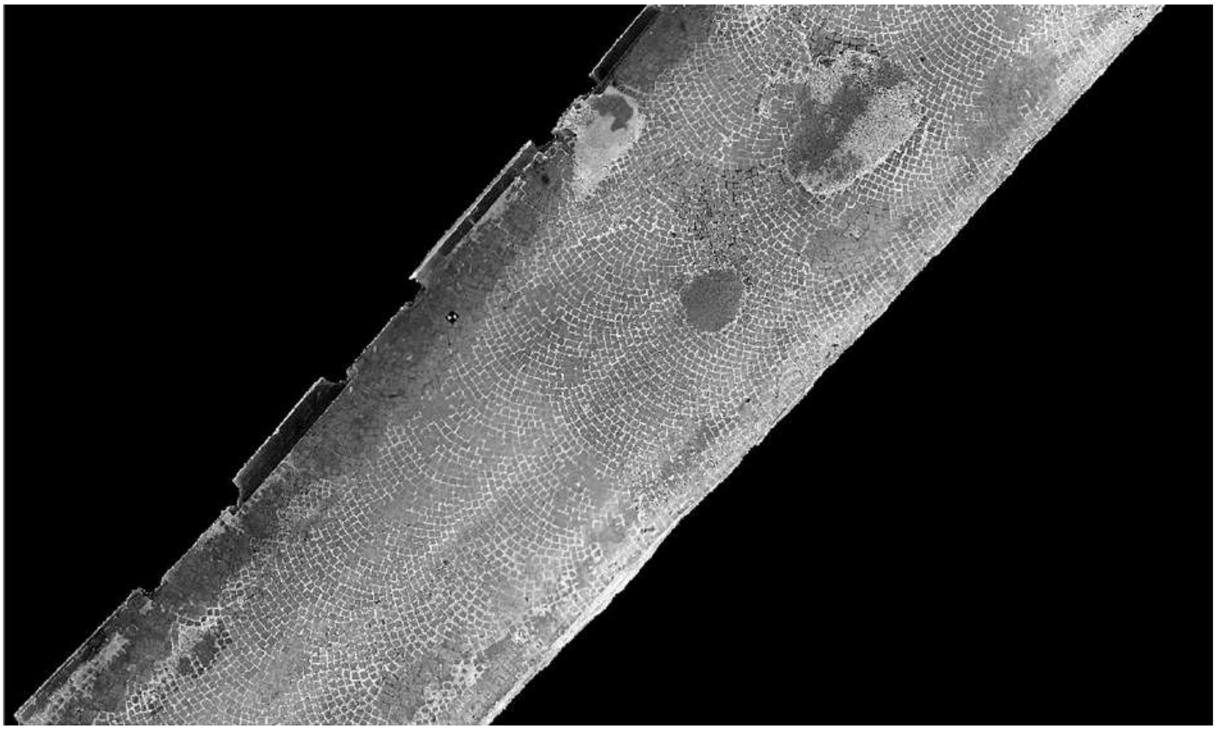


Figure 9 - PCA band 1 of a portion of pilot road section.

On the basis of the pre-processing operations carried out, the available bands were: R-G-B, H-S-V and PCA1- PCA2- PCA3. Initially PCA band 1, H band and V band were analyzed individually, then orthophotos in RGB and HSV space were considered. Finally, an attempt was made to integrate the various bands, which, although not increasing the available data, could have improved the automatic recognition of the laying pattern. All the settings were applied both to original and equalized images. Table 3 summarizes the band sets (BSs) considered in the tests.

Table 3 – Band sets used in the semi-automatic classification

Band set name	Number of bands	Used bands
BS1	1	PCA1
BS2	1	H
BS3	1	V
BS4	3	R, G and B
BS5	3	H, S and V
BS6	4	R, G, B and PCA1
BS7	4	R, G, B and H
BS8	7	R, G, B, H, S, V and PCA1

As far as the classification is concerned, for the first three BSs, which consist of a single band, the classification was made by setting a limit threshold that discriminates stone cubes and joints. For the other BSs, it was possible to use SCP automatic classification. SCP implements Land Cover Signature Classification (LCS) algorithm which allows the definition of spectral thresholds for each training input signature (a minimum value and a maximum value for each band). Spectral signatures of image pixels are compared to the training spectral signatures and a pixel belongs to a class if its spectral signature is completely contained in the spectral region defined by that class. Otherwise, if a pixel falls inside overlapping regions or outside any spectral region it will be not classified. To this issue LCS can be coupled with additional algorithms that determine how to classify ambiguous pixels. SCP implements three additional algorithms (Minimum Distance, Maximum Likelihood, Spectral Angle Mapping), which were all tested to achieve the best classification result. Minimum Distance (MD) algorithm calculates the Euclidean distance between spectral signatures of image pixels and training spectral signatures. Therefore, the distance is calculated for every pixel in the image, assigning the class of the spectral signature that is closer, according to a defined threshold. Maximum Likelihood (ML) algorithm calculates the probability distributions for the classes, estimating if a pixel belongs to a land cover class. In order to use this algorithm, a sufficient number of pixels is required for each training area allowing for the calculation of the covariance matrix. Finally, Spectral Angle Mapping (SA) algorithm calculates the spectral angle between spectral signatures of image pixels and training spectral signatures.

2.5.2. Neural Network

Most of the limitations of the previously described semi-automatic classification can be overcome by a convolutional neural network (CNN). Shadows and other lighting issues strongly affect the output of the previous technique since it is uniquely based on an evaluation of radiometric features on a per-pixel base. On the contrary, a CNN is capable of taking into consideration radiometric changes between adjacent pixels and of highlighting with (usually) greater flexibility features in a picture on the basis of its relevant shape, rather than solely on its color (or radiometric) features. In the experimentation, at this stage, the development of a new CNN architecture specifically tailored for classifying and detecting stone pavement pattern was considered unnecessary, since very reliable and consolidated solutions for pattern recognition and detection are already available. In particular, U-Net architecture [43], although being initially applied for biomedical tissue segmentation, has proven to be easily adaptable to a wide range of pattern segmentation problems and, at the time of writing, is probably the best performing architecture as far as the training input dataset is based on just few images, and localization (assigning every single pixel to a specific class/label) rather than classification is required. Differently from other approaches [44], where a sliding window is moved around the image and classified assigning the resulting class to the central pixel of the window itself for localization, in a U-Net CNN all the pixels of a tile of the image are classified adopting a two processing stages strictly concatenated: the first, that acts as an encoder, where the input image (tile) is subsequently simplified into a feature representation (through convolutions and max-pooling) and where the actual classification of information occurs, and the second, that acts as a decoder, where the discriminative features extracted in the previous stage are semantically projected onto the pixel space, upsampling (via transposed convolution) the condensed feature map up to the original resolution of the input image. In other words, during the encoding path, like in a traditional CNN, through the repeated application of convolution (and pooling) stages, the sparse information of the input image is condensed in downsampled feature maps. During the decoding (upsampling) stages, the feature maps expands and is concatenated with the correspondingly cropped feature map from the encoding path, in order to provide a classification on a per-pixel basis of the original input image.

As in any other classification procedure, using a neural network, the process consists of two phases: the first phase consists of network training, while the second phase consists of the prediction of objects in an image based on training data. Classification tests through the CNN have been carried out from the RGB orthophotos on which the already described manual classification has been made. From the distribution of material classes, it appeared that the data were imbalanced, i.e. there were too few examples of specific classes for training the CNN. Thus, all materials that do not constitute stone cubes were grouped into a single class resulting in a binary classification problem. The training dataset consists of areas of the source image (tiles), whose size (in pixels) is crucial for the success of the training and, consequently, of the classification, as in each of them is good to see an adequate number of items to be classified. An example is represented by the presence in each patch of a paved part (cubes) and a proportion of joints so that the CNN is able to get the greatest number of information from every patch. As a preliminary step different tile size were considered, trying to figure out the best size for training the network. Figure 10 shows examples of tiles with different dimensions (in pixels). Smaller tiles (32x32 pixel which is approximately the size of a single stone block at the actual orthophoto scale) allows, for the same training area, to obtain a greater number of training samples. However, in most cases, such a small tile does not provide enough significant content for the CNN to discriminate between the classes and might lead to unsatisfactory results. On the contrary, larger patches provide more context for better accuracy, but at the cost of leaving less available training samples [45]. Therefore, it was decided to perform three different analyses in which the orthophoto was portioned in patches each having dimensions of 256x256, 128x128 and 64x64 pixels, respectively.

Figure 10 – Patches of different size

A preliminary series of tests, whose results are not reported here for brevity, showed that data augmentation do not provide significant benefits (and in some circumstances actually lowered the CNN accuracy). It should be noted that, in this particular context, the size and orientation of the stone blocks are almost constant throughout the analyzed area: considering rotated or scaled training input, which usually should improve the generalization process of the neural network, in this case conducted to a lower performance of the system. Also, considering overlapping tiles, in order to increase the number of training data, have not improved significantly the final accuracy of the classification and simply made the training process lengthier. The neural network was implemented using Keras libraries and for each configuration set (with different tile size) four consecutive training cycles, each one considering 50 training epochs, were considered.

3. Results and discussions

3.1.Semi-Automatic Classification Plugin

The results obtained with the use of thresholds (BS1 to BS3) and supervised classification through SCP (BS4 to BS8) are listed in Table 4, while Figure 11 and Figure 12 show the graphical representation of the numerical evaluation results. For the supervised classification through SCP, only the results obtained from the combined use of LCS and the other classification algorithms (MD, ML, SA) are reported in Table 4 because, given the high overlap between the spectral signatures highlighted above, the simple use of LCS excluded most pixels from the classification. In the results evaluation, the focus will be mainly on accuracy values, F_1 score and MCC parameter.

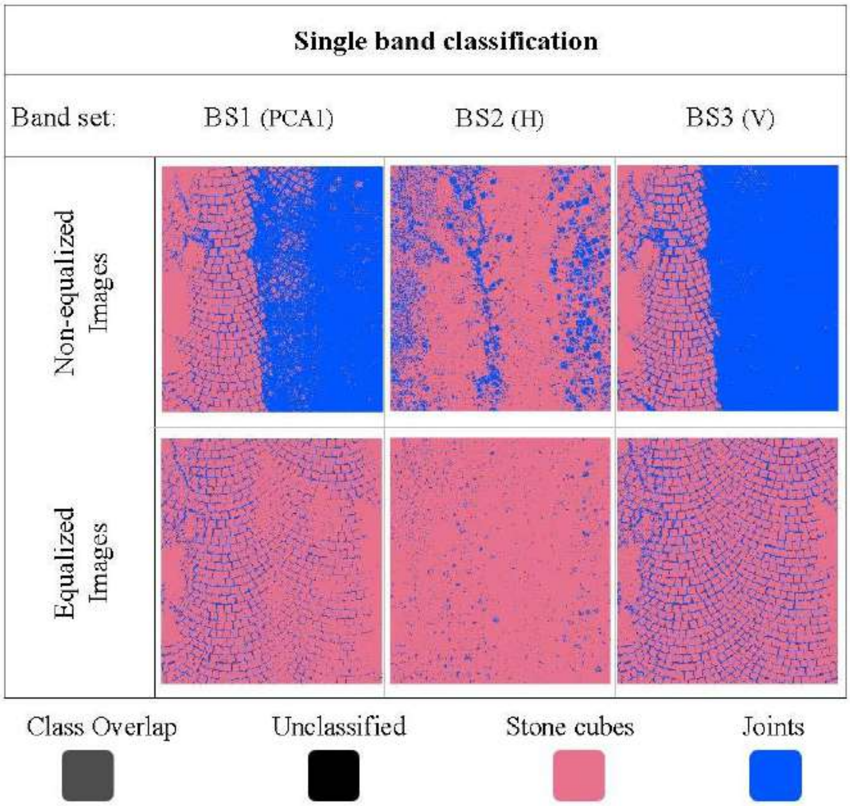


Figure 11 - Graphical representation of the numerical evaluation results of semi-automatic classification with single bands applied to non-equalized and equalized images.

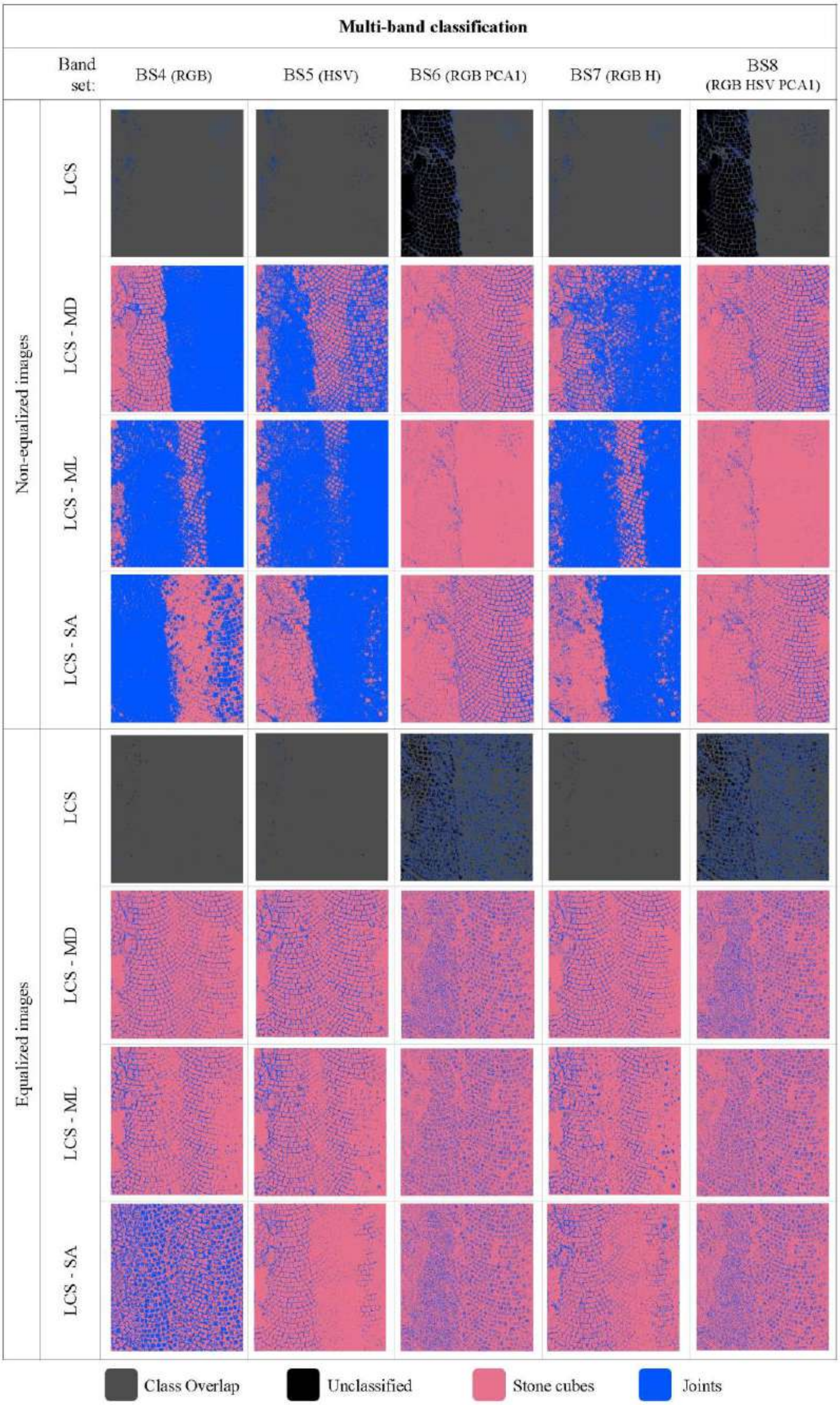


Figure 12 - Graphical representation of the numerical evaluation results of semi-automatic classification with different multi-band sets applied to non- equalized and equalized images.

Table 4 – Overall performances of semi-automatic classification with different band sets and classification algorithms

	Band set	Threshold/ Algorithm	Recall	Selectivity	Precision	Accuracy	F1	MCC
Non-equalized images	BS1 (PCA1)	285	0.464	0.914	0.960	0.547	0.625	0.301
	BS2 (H)	210	0.874	0.106	0.811	0.731	0.841	-0.025
	BS3 (V)	105	0.344	0.956	0.971	0.457	0.508	0.257
	BS4 (RGB)	LCS+MD	0.361	0.948	0.968	0.470	0.526	0.261
		LCS+ML	0.199	**0.991**	**0.990**	0.346	0.332	0.200
		LCS+SA	0.487	0.529	0.819	0.495	0.611	0.012
	BS5 (HSV)	LCS+MD	0.518	0.922	0.967	0.593	0.674	0.345
		LCS+ML	0.095	**0.991**	0.979	0.262	0.174	0.124
		LCS+SA	0.348	0.807	0.888	0.433	0.500	0.129
	BS6 (RGB-PCA1)	LCS+MD	0.963	0.560	0.906	0.888	0.934	0.598
		LCS+ML	**0.975**	0.022	0.814	0.798	0.887	-0.008
		LCS+SA	0.956	0.641	0.921	0.897	0.938	0.640
	BS7 (RGB H)	LCS+MD	0.353	0.970	0.981	0.467	0.519	0.276
		LCS+ML	0.260	0.986	0.988	0.395	0.412	0.233
		LCS+SA	0.339	0.694	0.829	0.405	0.481	0.027
	BS8 (RGB HSV PCA1)	LCS+MD	0.955	0.698	0.933	**0.907**	**0.943**	**0.681**
		LCS+ML	**0.975**	0.022	0.814	0.798	0.887	-0.008
		LCS+SA	0.968	0.513	0.897	0.883	0.931	0.572
Equalized images	BS1 (PCA1)	285	**0.993**	0.533	0.903	0.908	0.946	0.668
	BS2 (H)	210	0.974	0.013	0.812	0.795	0.886	-0.033
	BS3 (V)	105	0.983	**0.767**	**0.949**	**0.943**	**0.965**	**0.802**
	BS4 (RGB)	LCS+MD	0.991	0.617	0.919	0.921	0.953	0.721
		LCS+ML	0.975	0.563	0.907	0.899	0.940	0.633
		LCS+SA	0.579	0.420	0.814	0.550	0.677	-0.001
	BS5 (HSV)	LCS+MD	0.962	0.708	0.935	0.915	0.949	0.708
		LCS+ML	0.975	0.586	0.912	0.903	0.942	0.650
		LCS+SA	0.992	0.342	0.869	0.872	0.926	0.510
	BS6 (RGB-PCA1)	LCS+MD	0.768	0.081	0.786	0.640	0.777	-0.146
		LCS+ML	0.768	0.005	0.772	0.627	0.770	-0.225
		LCS+SA	0.768	0.101	0.789	0.644	0.778	-0.126
	BS7 (RGB H)	LCS+MD	0.988	0.653	0.926	0.926	0.956	0.740
		LCS+ML	0.954	0.468	0.887	0.864	0.920	0.498
		LCS+SA	0.966	0.430	0.881	0.866	0.922	0.496
	BS8 (RGB HSV PCA1)	LCS+MD	0.768	0.121	0.793	0.648	0.780	-0.105
		LCS+ML	0.768	0.005	0.772	0.627	0.770	-0.225
		LCS+SA	0.768	0.077	0.785	0.640	0.776	-0.150

As can be seen from Figure 11 and Figure 12, the best results are provided by using pre-equalized images where shadows have been mitigated. In the most of analyses performed on equalized images, the laying pattern is recognizable over the entire image. The V band (BS3), expressing the brightness relative to the same lighting conditions, is able to well highlight the intensity differences between stone cubes and joints and, using equalized image, provided the best statistical scores (accuracy 94.3%, F1 96.5 and MCC 80.2%). Instead, when using not equalized images, only the part in light (with illumination more similar to the training area) is correctly classified (Figure 11), reaching in that portion the 86.5% of accuracy, while globally the accuracy is 45.7%. The same behavior is given by band 1 (BS1) obtained from the principal component analysis, although with slightly lower accuracy than BS3. BS1 allows to correctly classify 90.8% of pixels in the pre-equalized images, while in the images with shadows the local accuracy in the areas in light is 90.8% and globally decreases to 54.7%. Unlike expected, the use of the H band alone does not give so good results. This may be due to the fact that, under reflections, the hues of stone cubes and joints are very similar. As far as the band sets with multiple bands are concerned, RGB band set (BS4) provides accuracies higher than 90% using pre-equalized images, while it is not able to classify correctly the image affected by shadows. The conversion to HSV space (BS5 and BS7) does not give very significant improvements compared to the use of RGB in equalized images. In non-equalized images, however, it does not overcome the shadow problem and the accuracies obtained are very low. On the other hand, the integration with the PCA1 component (BS6 and BS8), proves to be effective to un-correlate the parameters and overcome the problem of different lighting when applied to non-equalized images. In fact, the BS6 and BS8 are the ones that provide the best results, limited to non-equalized images, reaching accuracy percentages of about 90%. Looking at the graphical representations of the results in Figure 12, it can be seen that only these two band sets are not affected by the presence of shadows. In the other cases (in particular BS4 LCS+MD, BS4 LCS+SA, BS5 LCS+SA and BS7 LCS+SA) only the portion in light or shadow (alternatively) is correctly classified. In contrast, by applying BS6 and BS8 configurations to the equalized images, the results are reversed by inverting the pixels classified as stone cubes and as joints.

As for the algorithms used, considering the high overlap between spectral signatures, the simple use of LCS proved to be ineffective. For BS4, BS5 and BS7 almost all pixels fall into overlapping areas; for BS6 and BS8, when using non-equalized images, the areas in light are not classified and those in shadow are considered overlapping areas, while, when using equalized images, some classified pixels appear, but they are largely wrongly classified. Maximum Distance algorithm provides overall the best results and correctly classifies pixels with radiometric values falling in the overlapping zones between spectral signatures. Using non-equalized images, Maximum Likelihood algorithm does not correctly distinguish classes and classifies everything either as stone cubes or as joints. This algorithm provides the highest number of omissions or commissions, as demonstrated by the values of recall (the proportion of stone cubes that are correctly identified as such) and selectivity (the portion of pavement not made of stone cubes that is correctly identified as such) that have opposite values (high recall and low selectivity or vice versa). For instance, BS4 LCS+ML and BS5 LCS+ML have the highest values of precision, but very low recall score, which result in almost the entire area classified as joints while the stone cubes are not identified. For BS6 LCS+ML and BS8 LCS+ML the behavior is the opposite. With equalized images, instead, the difference between the algorithms is not so clear, although the best results are provided by Maximum Distance algorithm. Observing the other statistical parameters, both the F_1 score and the MCC parameter confirm the trend showed by accuracy and reach the best results using BS3 applied to equalized images.

3.2. Neural Network

The best performing training were obviously obtained considering a greater number of training epochs. The produced output during classification is a raster image, the same size as the original tested orthophoto, whose pixels have a probability value associated, ranging between 0.0 and 1.0, that represents the likelihood, according to the CNN network, the element is respectively a joint or a stone cube. In other words, a value of 0.0 means that, according the CNN classifier, the pixel should be classified as a joint, while a value of 1.0 represent, most likely, a stone cube. Since every pixel might have an intermediate value between 0 and 1 a constant threshold value set to be 0.5 was considered for discriminate between the two classes. Some authors [46] suggest to consider, during the CNN training stage, also the threshold as hyper-parameter to further improve the network performance. However, in our tests, such additional optimization was not required since the two classes were strongly separated at the end of the prediction process. As can be seen in Table 5, the best accuracy is obtained (as the reader can easily guess) at the fourth training cycle (i.e. after 200 training epochs) regardless the patch size. However, analyzing intermediate results, i.e. the performance of the classifier with less training epochs (every training cycle added 50 training epochs in the procedure so, after training cycle 1, 2, 3 and 4 the U-Net network was trained considering respectively 50, 100, 150 and 200 epochs) can be interesting to highlight a faster or slower performance of the network toward optimal results. As can be seen in Table 5, in all cases the accuracy is higher than 0.90 and increases as the tile size decreases. This indicates that the 64x64 pixel patches (which correspond to areas of 128x128 mm) are large enough to cover a larger surface area and at the same time contain sufficient information about both joints and stone cubes. Analyzing the recall and the selectivity, the best results are obtained with 128x128 pixels patches. The same result is obtained also for the precision. In statistical analysis of binary classification, the F_1 score, which is a measure of a test's accuracy, makes it possible to consider simultaneously the precision and the recall. Considering this parameter, the best performance is obtained at the fourth training cycle for each patch size: in all cases this parameter is higher than 0.94 and increases as the patch size decreases. However, all the considered parameters are representative of the analysis only if the analyzed classes have similar dimensions, i.e. only if they are balanced. In this case the pixel labelled as stone cubes are about four times than those labelled as joints. For this reason, it was decided to analyze the MCC parameter. Also in this case the best performance is obtained at the fourth training cycle for each patch size and increase as the patch size decrease. This indicates that the 64x64 pixel patches contain sufficient information about both joints and stone cubes. However, although this analysis has shown good results on stone pavement made of 6/8 class stone cubes, patches of such sizes may not be suitable for the analysis of stone pavement made of elements of different sizes. For example, with orthophoto of equal resolution, if 10/12 class stone cubes were used 64x64 pixel patches might not be sufficient to contain enough information about both joints and stone cubes. Thus, the choice of the patches size should be deepened by varying the stone elements size used.

Table 5 - Overall performances of CNN classification with different patches size and training cycles

Patches size	Training cycle	Accuracy	Recall	Selectivity	Precision	F1	MCC
256x256	1	0.840	0.960	0.352	0.858	0.906	0.411
	2	0.832	0.857	0.731	0.928	0.891	0.535
	3	0.804	0.820	0.739	0.927	0.870	0.490
	4	0.908	0.980	0.614	0.912	0.945	0.687
128x128	1	0.827	**0.991**	0.162	0.828	0.902	0.313
	2	0.846	0.817	**0.967**	0.990	0.895	0.659
	3	0.925	0.932	0.900	0.974	0.952	0.783
	4	0.942	0.942	0.938	**0.984**	0.963	0.831
64x64	1	0.950	0.966	0.881	0.971	0.969	0.842
	2	0.948	0.957	0.910	0.977	0.967	0.842
	3	0.933	0.937	0.916	0.979	0.957	0.805
	4	**0.954**	0.967	0.899	0.975	**0.971**	**0.856**

Figure 13 shows the graphical representation of the numerical evaluation results. As can be seen, when 256x256 pixel patches are used in the checks carried out for all four training cycles, CNN is not able to adequately distinguish the two classification elements: in particular, there is a difficulty in recognizing the stone cubes in sunny areas (256x256 patch size in training cycle 2 and 3). As far as 128x128 patches are concerned, already from the second training cycle the CNN has been able to distinguish adequately the joints from the stone cubes. Finally, considering the training done with the 64x64 pixel patches (ca. twice the size of a single stone block), a very good result was obtained already after the first training cycle.

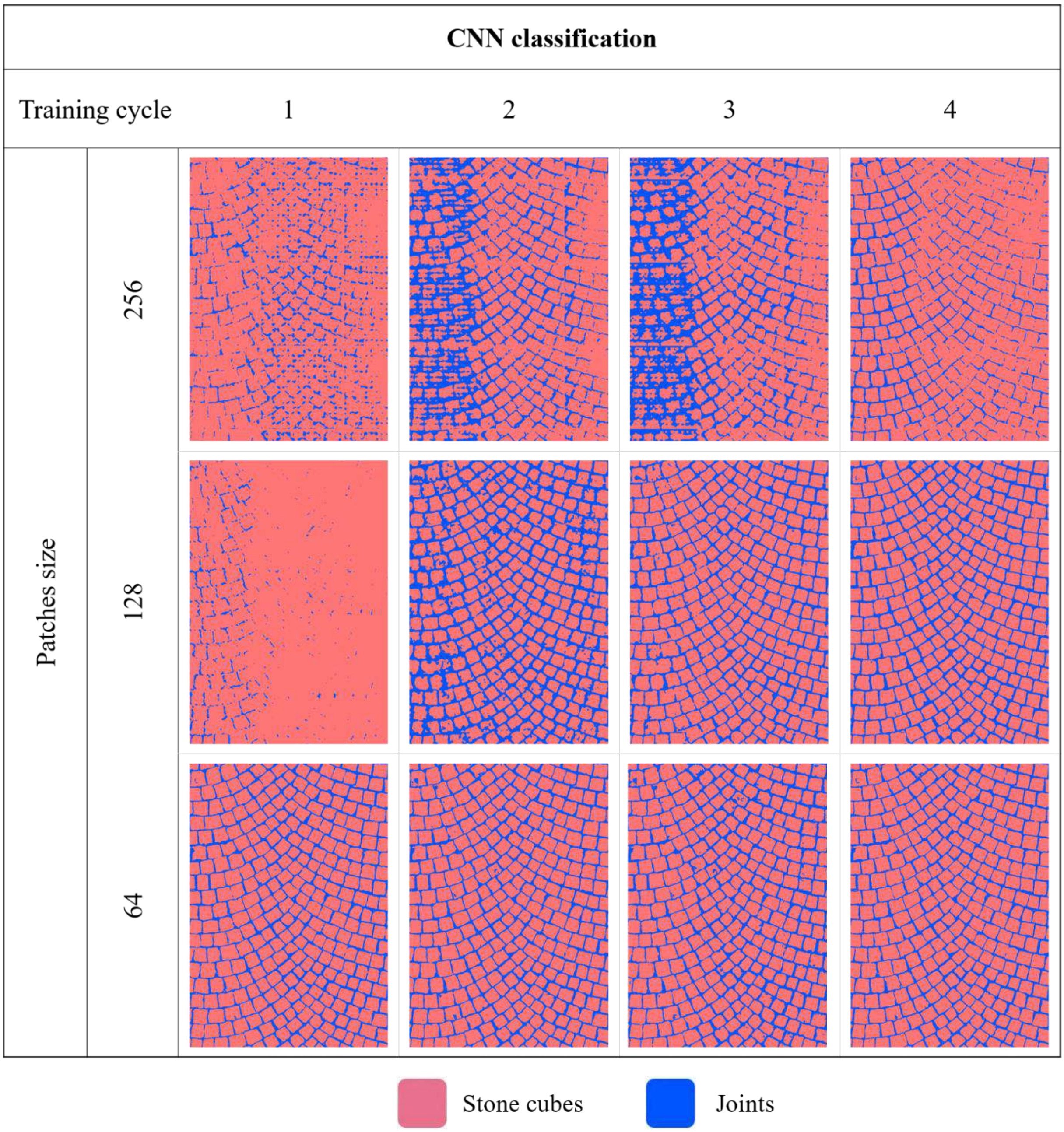


Figure 13 – Graphical representation of the numerical evaluation results of CNN classification with different patches size and training cycles.

Considering the results obtained after the fourth training cycle with the 64x64 pixels patches, and comparing the values of the coefficients shown in the table with the graphical representation of the numerical evaluation results, it can be noted that despite the number of pixels recognized as belonging to the “joints” class is less than 90% of the pixels actually belonging to that class, the location of the element “joint” inside the pavement is correct. For that reason, the automatic classification using a convolutional neural network on images obtained from unmanned aerial vehicle has proven to be an excellent alternative to the traditional manual inspection and can be implemented for other types of stone block pavements, also with the aim of distress identification.

4. Conclusions

This study investigated the possibility of using different methodologies for the automatic detection of a stone pavement's pattern based on UAV photogrammetry and the possibility of inserting them into an urban pavement management system. The analysis was carried out on stone pavement consisting of small, cubic, Trentino's porphyric elements, commercially in 6/8 class placed in an overlapping arcs laying pattern. For the automatic detection, two approaches were used: supervised classification through semi-automatic classification plugin (SCP) and convolutional neural network (CNN). The SCP was applied to eight different band sets, combination of the 7 available bands (R, G, B, PCAband1, H, S and V), with or without radiometric equalization to reduce different illumination condition of the tested scenes (e.g. areas with shadows vs areas with direct sun light exposure), and different algorithms (land cover signature, minimum distance, maximum likelihood and spectral angle mapping). Convolutional neural network was tested with patches of different size (256x256, 128x128 and 64x64 pixels) and four consecutive training cycles.

Based on these investigations, the following conclusions can be made:

- The results obtained with SCP have shown that the best accuracy, as well as the best MCC, were provided applying the classification process to pre-equalized images, where lighting conditions are balanced with the reference image used for training and where sharp shadows are mitigated. The locally adaptive filter used in these tests has proven to be effective, but its performance is strongly influenced by the sliding window size used. In addition, since the algorithm levels the RGB values to match a reference image, the equalization is only effective if the whole scene consists of elements with the same characteristics as the reference image (e.g. only stone pavements). On the contrary, if extraneous elements are visible (e.g. patches of incongruous materials, fences etc.) the algorithm tends to equalize even the RGB values of those elements, introducing ambiguities in the final classification. This issue could be fixed using more sophisticated shadow removal algorithms, which, however, would probably require a more demanding parameter tuning.

- Performing a principal component analysis of RGB images was useful for un-correlating parameters and maximize the differences between different types of pavements regarding illumination conditions.
- The best accuracy, as well as the best MCC, for the supervised classification through SCP was obtained for the BS3 (V band only) applied to equalized images. In this type of pavements, the V band, representing the brightness relative to the same lighting conditions, is able to highlight more the intensity differences between stone cubes and joints.
- The results obtained with the U-Net CNN have shown that the best accuracy and MCC value were obtained with a tile size which is approximately twice the size of a single stone block (64 x 64 pixel). In this case the CNN reached accuracy and MCC values greater than 0.95 and 0.85 respectively. At the same time, the same 64x64 pixel tile size, allowed obtaining very good results also at the end of the first (50 epochs only) training cycle.
- However, although this analysis has shown good results on stone pavement made of 6/8 class stone cubes, patches of such sizes may not be suitable for the analysis of stone pavement made of elements of different sizes or joints with different width. For

this reason, the choice of the patches size should be considered carefully, especially if varying stone or joint elements are present.

- U-Net CNN classification has proven not to be affected by the influence of lighting conditions and shadows, so it does not require image pre-processing through equalization, principal component analysis or conversion to HSV space. In this way, potential alterations of the original data due to the pre-processing phase are excluded.

In light of the above, CNN classification, as far as these experiences are considered, proved to be a more flexible and efficient approach: it surely requires a carefully planned and probably more time consuming training stage but, in the end, provides not only better results but also, and more importantly, a higher level of reliability. To obtain comparable results with the other technique (e.g. accuracy of 94.3% obtained with V equalized band analysis versus a CNN accuracy of 95.4%) the user should tune a lot of different parameters and choose carefully the best (more representative) image band. On the contrary CNN, even with the use of less efficient patch sizes and without any equalization or image pre-processing stage, provides quite satisfactory results. In conclusion, the automatic classification using a U-Net CNN on images obtained from UAV has proven to be an excellent alternative to the traditional manual inspection and can be implemented for other types of stone pavements, also with the aim of distress identification.

Acknowledgments

This research was developed within the Projects of National Interest – PRIN 2017 "Stone pavements. History, conservation, valorisation and design" (20174JW7ZL) financed by the Ministry of Education, University and Research (MIUR) of the Italian Government.

References

[1] M.R. Jelokhani-Niaraki, A.A. Alesheikh, A. Alimohammadi, A. Sadeghi-Niaraki, K. Kim, An approach for automatic updating of GIS road segments for a pavement management system (PMS), Journal of Spatial Science. 56 (2011) pp. 253–267. https://doi.org/10.1080/14498596.2011.623346.

[2] V. Donev, M. Hoffmann, Optimisation of pavement maintenance and rehabilitation activities, timing and work zones for short survey sections and multiple distress types, International Journal of Pavement Engineering. 21 (2020) pp. 583–607. https://doi.org/10.1080/10298436.2018.1502433.

[3] K.A. Abaza, S.A. Ashur, I.A. Al-Khatib, Integrated pavement management system with a Markovian prediction model, Journal of Transportation Engineering. 130 (2004) pp. 24–33. https://doi.org/10.1061/(ASCE)0733-947X(2004)130:1(24).

[4] G. Loprencipe, A. Pantuso, P. Di Mascio, Sustainable Pavement Management System in Urban Areas Considering the Vehicle Operating Costs, Sustainability (Switzerland). 9 (2017). https://doi.org/10.3390/su9030453.

[5] M.V. Corazza, P. Di Mascio, L. Moretti, Managing sidewalk pavement maintenance: A case study to increase pedestrian safety, Journal of Traffic and Transportation Engineering (English Edition). 3 (2016) pp. 203–214. https://doi.org/10.1016/j.jtte.2016.04.001.

[6] W.D. Cottrell, S. Bryan, B.R. Chilukuri, V. Kalyani, A. Stevanovic, J. Wu, Transportation Infrastructure Maintenance Management: Case Study of a Small Urban City, Journal of Infrastructure Systems. 15 (2009) pp. 120–132. https://doi.org/10.1061/(ASCE)1076-0342(2009)15:2(120).

[7] A. Osorio, A. Chamorro, S. Tighe, C. Videla, Calibration and Validation of Condition Indicator for Managing Urban Pavement Networks, Transportation Research Record:

Journal of the Transportation Research Board. 2455 (2014) pp. 28–36. https://doi.org/10.3141/2455-04.

[8] Y.U. Shah, S.S. Jain, M. Parida, Evaluation of prioritization methods for effective pavement maintenance of urban roads, International Journal of Pavement Engineering. 15 (2014) pp. 238–250. https://doi.org/10.1080/10298436.2012.657798.

[9] M.Y. Shahin, Pavement management for airports, roads, and parking lots: Second edition, 2005, ISBN 9780387234649. https://doi.org/10.1007/b101538.

[10] P. Zoccali, G. Loprencipe, A. Galoni, Sampietrini stone pavements: distress analysis using pavement condition index method, Applied Sciences. 7 (2017). https://doi.org/10.3390/app7070669.

[11] E. Garilli, F. Autelitano, F. Giuliani, A study for the understanding of the Roman pavement design criteria, Journal of Cultural Heritage. 25 (2017) pp. pp.87-93. https://doi.org/10.1016/j.culher.2017.01.002.

[12] E. Garilli, F. Giuliani, Stone pavement materials and construction methods in Europe and North America between the 19th and 20th century, International Journal of Architectural Heritage. 13 (2019) pp. 742–768. https://doi.org/10.1080/15583058.2018.1470269.

[13] F. Autelitano, E. Garilli, F. Giuliani, Criteria for the selection and design of joints for street pavements in natural stone, Construction and Building Materials. 259 (2020). https://doi.org/10.1016/j.conbuildmat.2020.119722.

[14] F. Dutruel, J. Dardare, Contribution to the study of structural behaviour of a concrete block pavement., in: Proceeding of Second International Conference on Concrete Block Paving, Delft, 1984: pp. 29–39.

[15] E. Garilli, F. Autelitano, R. Roncella, F. Giuliani, The influence of laying patterns on the behaviour of historic stone pavements subjected to horizontal loads, Construction

and Building Materials. 258 (2020). https://doi.org/10.1016/j.conbuildmat.2020.119657.

[16] Associates Applied Research, Interlocking concrete pavement distress manual: tools for condition assessment, performance modeling and pavement management for a long service life, Toronto, 2007.

[17] T.B.J. Coenen, A. Golroo, A review on automated pavement distress detection methods, Cogent Engineering. 4 (2017). https://doi.org/10.1080/23311916.2017.1374822.

[18] Y. Tan, Y. Li, UAV photogrammetry-based 3D road distress detection, ISPRS International Journal of Geo-Information. 8 (2019). https://doi.org/10.3390/ijgi8090409.

[19] J. Landa, D. Prochazka, Automatic Road Inventory Using LiDAR, Procedia Economics and Finance. 12 (2014) pp. 363–370. https://doi.org/10.1016/s2212-5671(14)00356-6.

[20] A. Mancini, E.S. Malinverni, E. Frontoni, P. Zingaretti, Road pavement crack automatic detection by MMS images, in: 2013 21st Mediterranean Conference on Control and Automation, MED 2013 - Conference Proceedings, 2013: pp. 1589–1596. https://doi.org/10.1109/MED.2013.6608934.

[21] S. Mathavan, M. Rahman, K. Kamal, Use of a Self-Organizing Map for Crack Detection in Highly Textured Pavement Images, Journal of Infrastructure Systems. 21 (2015). https://doi.org/10.1061/(ASCE)IS.1943-555X.0000237.

[22] S. Mathavan, M.M. Rahman, M. Stonecliffe-Janes, K. Kamal, Pavement raveling detection and measurement from synchronized intensity and range images, 2014, ISBN 9780309295444. https://doi.org/10.3141/2457-01.

[23] C. Koch, I. Brilakis, Pothole detection in asphalt pavement images, Advanced Engineering Informatics. 25 (2011) pp. 507–515.

https://doi.org/10.1016/j.aei.2011.01.002.

[24] L. Huidrom, L.K. Das, S.K. Sud, Method for Automated Assessment of Potholes, Cracks and Patches from Road Surface Video Clips, Procedia - Social and Behavioral Sciences. 104 (2013) pp. 312–321. https://doi.org/10.1016/j.sbspro.2013.11.124.

[25] H. Maeda, Y. Sekimoto, T. Seto, T. Kashiyama, H. Omata, Road Damage Detection and Classification Using Deep Neural Networks with Smartphone Images, Computer-Aided Civil and Infrastructure Engineering. 33 (2018) pp. 1127–1141. https://doi.org/10.1111/mice.12387.

[26] S. Agnisarman, S. Lopes, K. Chalil Madathil, K. Piratla, A. Gramopadhye, A survey of automation-enabled human-in-the-loop systems for infrastructure visual inspection, Automation in Construction. 97 (2019) pp. 52–76. https://doi.org/10.1016/j.autcon.2018.10.019.

[27] S. Jiang, J. Zhang, Real-time crack assessment using deep neural networks with wall-climbing unmanned aerial system, Computer-Aided Civil and Infrastructure Engineering. 35 (2020) pp. 549–564. https://doi.org/10.1111/mice.12519.

[28] J.M. Vazquez-Nicolas, E. Zamora, I. González-Hernández, R. Lozano, H. Sossa, PD+SMC Quadrotor Control for Altitude and Crack Recognition Using Deep Learning, International Journal of Control, Automation and Systems. 18 (2020) pp. 834–844. https://doi.org/10.1007/s12555-018-0852-9.

[29] T. Ghosh Mondal, M.R. Jahanshahi, R.-T. Wu, Z.Y. Wu, Deep learning-based multi-class damage detection for autonomous post-disaster reconnaissance, Structural Control and Health Monitoring. 27 (2020) pp. 1–7. https://doi.org/10.1002/stc.2507.

[30] W. Wu, M.A. Qurishee, J. Owino, I. Fomunung, M. Onyango, B. Atolagbe, Coupling Deep Learning and UAV for Infrastructure Condition Assessment Automation, in: 2018 IEEE International Smart Cities Conference, ISC2 2018, 2019.

https://doi.org/10.1109/ISC2.2018.8656971.

[31] L. Congedo, Semi-Automatic Classification Plugin Documentation, (n.d.). https://semiautomaticclassificationmanual-v5.readthedocs.io/it/latest/# (accessed July 29, 2020).

[32] Qgis.org, Welcome to the QGIS project!, (n.d.). https://qgis.org/en/site/ (accessed July 29, 2020).

[33] J.C. Leachtenauer, R.G. Driggers, Surveillance and Reconnaissance Imaging Systems-Modeling and Performance Prediction, Artch House, Boston, MA, USA, 2001, ISBN 978-1630812331.

[34] K. Kraus, I.A. Harley, S. Kyle, Photogrammetry, De Gruyter, 2007, ISBN 9783110892871. https://doi.org/10.1515/9783110892871.

[35] V. Ihlen, Landsat 7 (L7) Data Users Handbook, Sioux Falls, South Dakota, USA, 2019.

[36] V. Jain, A. Khunteta, Shadow removal for umbrageous information recovery in aerial images, in: 2017 International Conference on Computer, Communications and Electronics, COMPTELIX 2017, 2017: pp. 536–540. https://doi.org/10.1109/COMPTELIX.2017.8004028.

[37] S. Luo, H. Shen, H. Li, Y. Chen, Shadow removal based on separated illumination correction for urban aerial remote sensing images, Signal Processing. 165 (2019) pp. 197–208. https://doi.org/10.1016/j.sigpro.2019.06.039.

[38] N. Mo, R. Zhu, L. Yan, Z. Zhao, Deshadowing of Urban Airborne Imagery Based on Object-Oriented Automatic Shadow Detection and Regional Matching Compensation, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. 11 (2018) pp. 585–605. https://doi.org/10.1109/JSTARS.2017.2787116.

[39] P.M. Dare, Shadow analysis in high-resolution satellite imagery of urban areas, Photogrammetric Engineering and Remote Sensing. 71 (2005) pp. 169–177.

https://doi.org/10.14358/PERS.71.2.169.

[40] A.R. Smith, Color gamut transform pairs, in: Proceedings of the 5th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH 1978, 1978: pp. 12–19. https://doi.org/10.1145/800248.807361.

[41] E. Grilli, F. Remondino, Classification of 3D digital heritage, Remote Sensing. 11 (2019). https://doi.org/10.3390/RS11070847.

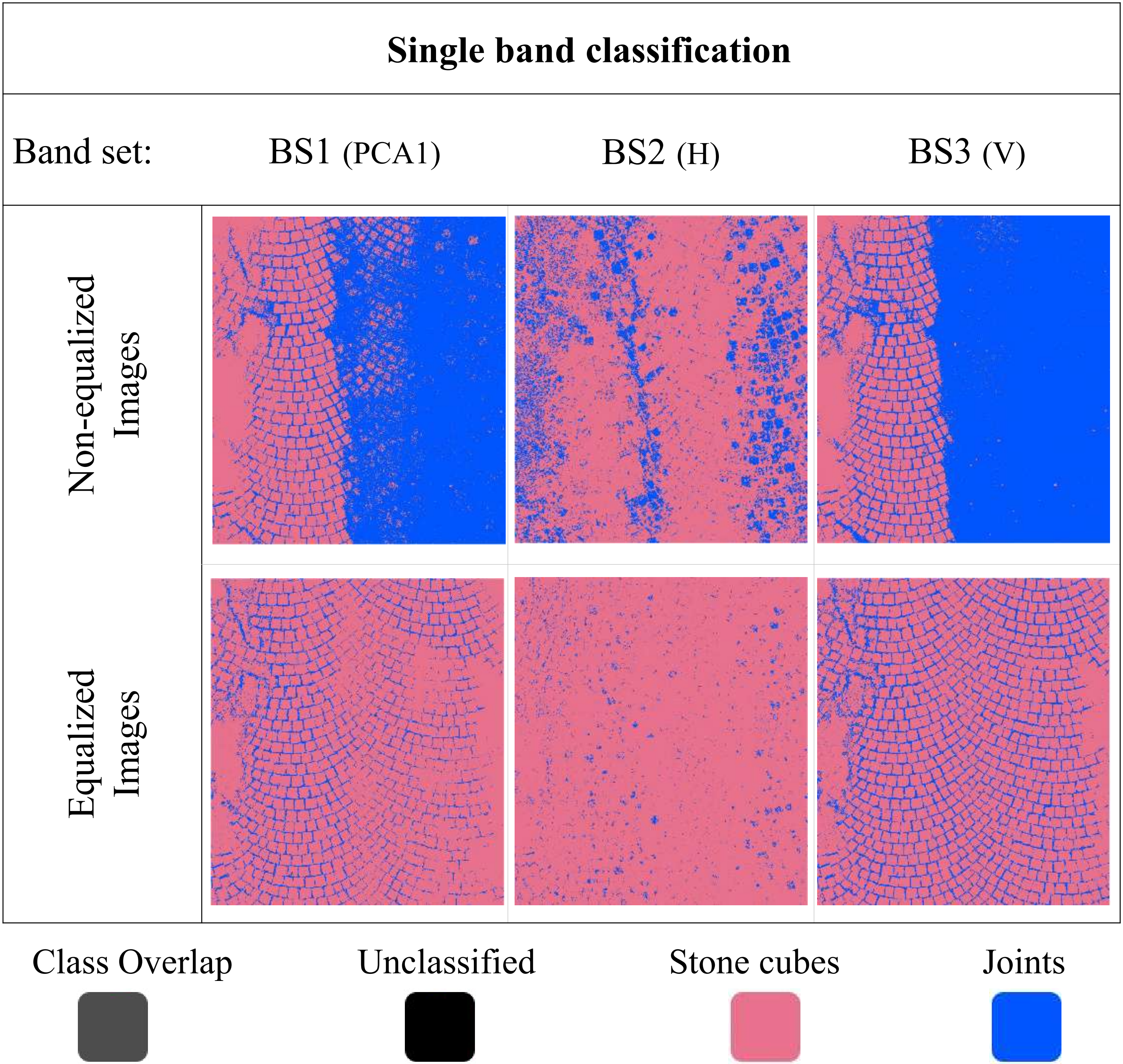
[42] I.T. Jolliffe, Principal Component Analysis, Springer Verlag, New york, 1986, ISBN 978-0-387-95442-4. https://doi.org/doi:10.1007/b98835.

[43] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, 2015, ISBN 9783319245737. https://doi.org/10.1007/978-3-319-24574-4_28.

[44] D.C. Cireşan, A. Giusti, L.M. Gambardella, J. Schmidhuber, Deep neural networks segment neuronal membranes in electron microscopy images, in: Advances in Neural Information Processing Systems, 2012: pp. 2843–2851.

[45] A. Riid, R. Lõuk, R. Pihlak, A. Tepljakov, K. Vassiljeva, Pavement distress detection with deep learning using the orthoframes acquired by a mobile mapping system, Applied Sciences (Switzerland). 9 (2019). https://doi.org/10.3390/app9224829.

[46] T. Fawcett, ROC graphs: Notes and practical considerations for researchers, Machine Learning. 31 (2004) pp. 1–38.


Single band classification
Band set:
BS1 (PCA1)
BS2 (H)
BS3 (V)
Non-equalized Images
Equalized Images
Class Overlap
Unclassified
Stone cubes
Joints

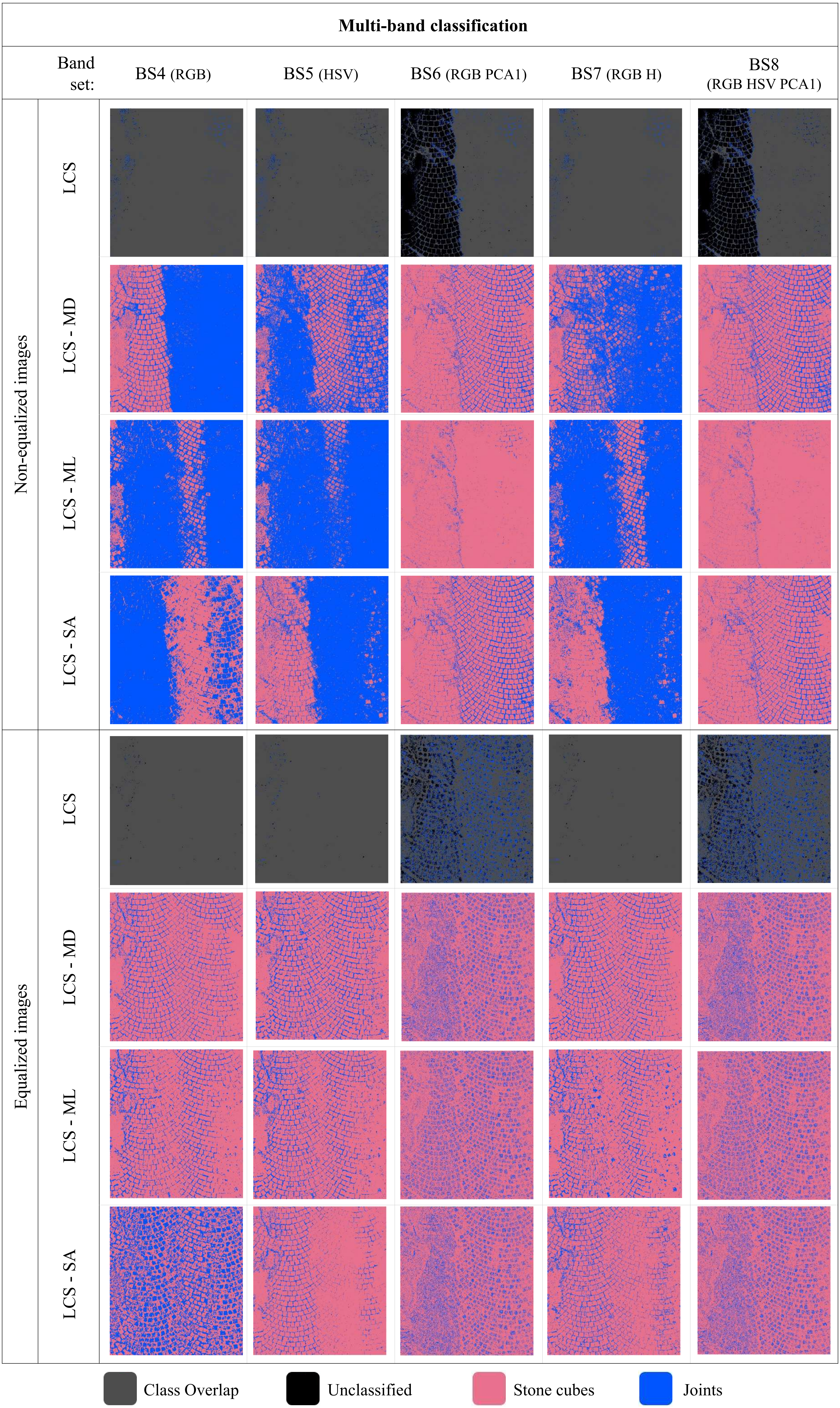
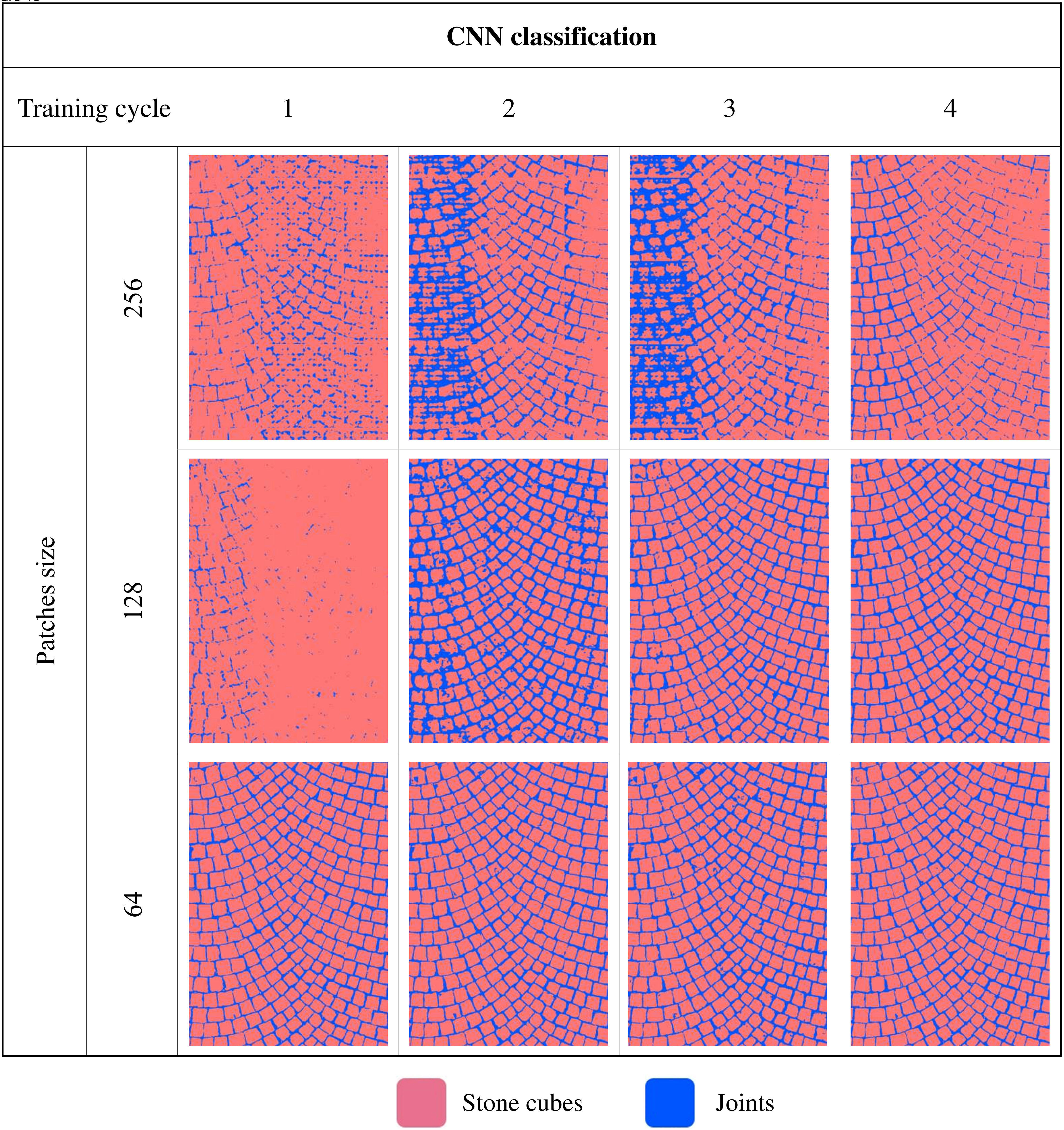

Multi-band classification
Band set:
BS4 (RGB)
BS5 (HSV)
BS6 (RGB PCA1)
BS7 (RGB H)
BS8 (RGB HSV PCA1)
Non-equalized images
Equalized images
LCS
LCS - MD
LCS - ML
LCS - SA
Class Overlap
Unclassified
Stone cubes
Joints

Figure 13


CNN classification
Training cycle
1
2
3
4
Patches size
256
128
64
Stone cubes
Joints