Ref. Ares(2021)3874058 - 14/06/2021

D6.3 Methodology for routine CH monitoring with multi-type remote sensing

Deliverable number	**D6.3**
Deliverable title	**Methodology for routine CH monitoring with multi-type remote sensing**
Nature[1]	**R**
Dissemination Level [2]	**PU**
Authors (email) Institution	Karathanassi V. karathan@survey.ntua.gr, Georgopoulos A. drag@central.ntua.gr, Kolokoussis P. pol@survey.ntua.gr Skamantzari M. sk.margarita@hotmail.com Karamvassis Kl. Karamvasis_K@hotmail.com Kristollari V. vkristoll@gmail.com Agrafiotis P. agrafiotis.panagiotis@gmail.com **NTUA** Claudio Mazzoli claudio.mazzoli@unipd.it **INIPD**
Editor (email) Institution	Karathanassi V. karathan@survey.ntua.gr **NTUA**
Leading partner	**NTUA**
Participating partners	**INIPD, IUAV**

[1] **R**=Document, report; **DEM**=Demonstrator, pilot, prototype; **DEC**=website, patent fillings, videos, etc.; **OTHER**=other

[2] **PU**=Public, **CO**=Confidential, only for members of the consortium (including the Commission Services), **CI**=Classified, as referred to in Commission Decision 2001/844/EC

Official submission date:	**31/05/2021**
Actual submission date:	

Modifications Index	
Date	Version
27/05/2021	1.0 First version
07/06/2021	1.1 Quality Review
08/06/21	2.0 Version

This work is a part of the HYPERION project. HYPERION has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no 821054.

Content reflects only the authors' view and European Commission is not responsible for any use that may be made of the information it contains.

ACRONYMS AND ABBREVIATIONS

CH	Cultural Heritage
HS	Hyperspectral
GGS	Generic Ground Station
UASSAR	Synthetic Aperture Radar
UAS	Unmanned Aircraft Systems
RS-MMS	Remote Sensing-based Multiscale Monitoring System
HRAP	Holistic Resilience Assessment Platform
SL	Spectral Library
CNN	Convolutional Neural Networks
SG	Structural-Geotechnical

Table of Contents

List of Figures

List of Tables

Executive Summary

The deliverable 6.3 focuses on the description of methods and tools that have been developed for the routine monitoring of the CH assets and broad areas as well, in the framework of the HYPERION project. The developed methods, apart from those related to deterioration mapping and material loss estimation, were applied on all the pilot sites (city of Rhodes, Granada, Venice and Tønsberg) and CH assets, and their products serve as reference data for the vulnerability models and post disaster damage assessment.

In chapter 2, satellite remote sensing methods developed within the HYPERION project are documented. Advanced methodologies using PS and SBAS functionalities on Sentinel-1 time-series satellite data have been developed for producing long-term ground deformation maps. Additionally, several Convolutional Neural Network architectures have been applied on very high-resolution Quickbird and WV2/3 images and evaluated for their ability to produce reliable land cover change detection maps with emphasis on the increase of the impervious materials.

In chapter 3, 3D modelling methods and analysis techniques were explored for the CH assets of the four pilot sites. Data acquired by cameras mounted on drones and mobile laser scanners have been used for the 3D representation of the CH assets. The produced point clouds, light models, texture models and sections provide a detailed 3D documentation for the assets and serve as input data for the SG simulator that is developed within the HYPERION project. Additionally, they serve as the basis for material loss estimation and will be used as reference data for creating 4D representations of important infrastructures over different time instances and detect structural deformations and other types of alterations.

Finally, in chapter 4, an integrated methodology based on 3D models and hyperspectral data has been developed for providing deterioration and material loss maps. Data acquired by mobile hyperspectral sensors were used in order to build a Spectral Library for the CH asset materials, identify their deviations from normal situations and produce deterioration maps. SAGA-GIS functionalities have been used to provide material loss maps. Additionally, deep learning methods have been used to detect flaws and defects such as cracks on the infrastructure by analysing the data received.

The deliverable is intended for public use, and it will particularly be helpful for the partners involved in the design of the HYPERION platform.

1. Introduction

1.1 Background

The deliverable D6.3 “Methodology for routine CH monitoring with multi-type remote sensing” focuses on the routine monitoring of the CH assets and the broad area as well, for hazards with slow or gradual onset. The baselines for routine monitoring were set in D6.1 document as a part of the remote sensing-based multiscale monitoring system (RS-MMS) operational functions. In the same document, functions for the routine monitoring operation of the RS-MMS were defined and necessary data and state of the art methods for their deployment were reported. In this document, the specific methods and tools that have been developed within the HYPERION project for the routine monitoring are being reported and the routine monitoring products are presented.

1.2 Scope and Objective

The objective of this deliverable is the description of the methods that have been developed for routine monitoring function of the RS-MMS. Routine monitoring includes displacement and land cover change detection maps of the broad area for all the pilot sites (city of Rhodes, Granada, Venice and Tønsberg), three-dimensional models for all the CH assets that are studied within the Hyperion project, and deterioration and material loss estimation for specific parts of the facades of the Fort of Saint Nikolas and the Roman bridge in the Rhodes island.

Satellite data after appropriate processing provide significant “background” information by pointing out hazards with a slow or gradual onset in the broad area of the CH monuments and facilitating change assessment including ground deformation and land cover changes. Since monuments are not isolated in the geographical space, assessment and evaluation of changes in their broader area are important because they serve as warning signals to the concerned stakeholders and facilitate them to take measures for preventing CH asset damages. To this end, advanced methodologies using PS and SBAS functionalities and Convolutional Neural Network architectures have been applied on satellite data aiming to produce reliable land deformation and land cover change detection maps, respectively.

Another main objective of the HYPERION project is to provide 3D models of the CH assets. Data obtained by the Generic Ground Station have been properly processed for producing 3D representations. Specific goals of this task are the development of point clouds, light models, texture models and sections, in order to obtain a detailed 3D representation of the assets, which will promote SG simulator functions, material loss estimation, and 4D representations over different time instances.

Other objectives of HYPERION project are to explore hyperspectral processing methods for fast and efficient assessment of the material deterioration level. Hyperspectral data provided by the GGS were used in order to build a Spectral Library for CH asset materials, and spectral unmixing methods have been applied to identify

spectral deviations from normal situations and produce material deterioration maps. Additionally, deep learning methods have been used to detect flaws and defects on the monuments. Finally, specific terrain analysis techniques of SAGA-GIS have been applied on the 3D models for estimating material loss. Integration of the 3D model based material loss estimation with hyperspectral imagery spectral unmixing results provide complete assessment of the CH material degradation status.

2. Study of the broad area for hazards with a slow or gradual onset

2.1 Ground displacement

Ground displacement monitoring is a critical factor to improve the awareness and help early recognition of potential risks. Understanding the ground deformation process is essential for better planning of CH risk mitigation actions. Ground deformation results were produced for all four pilot sites (Figure 1).

Figure 1: The four pilot sites

2.1.1 Tonsberg

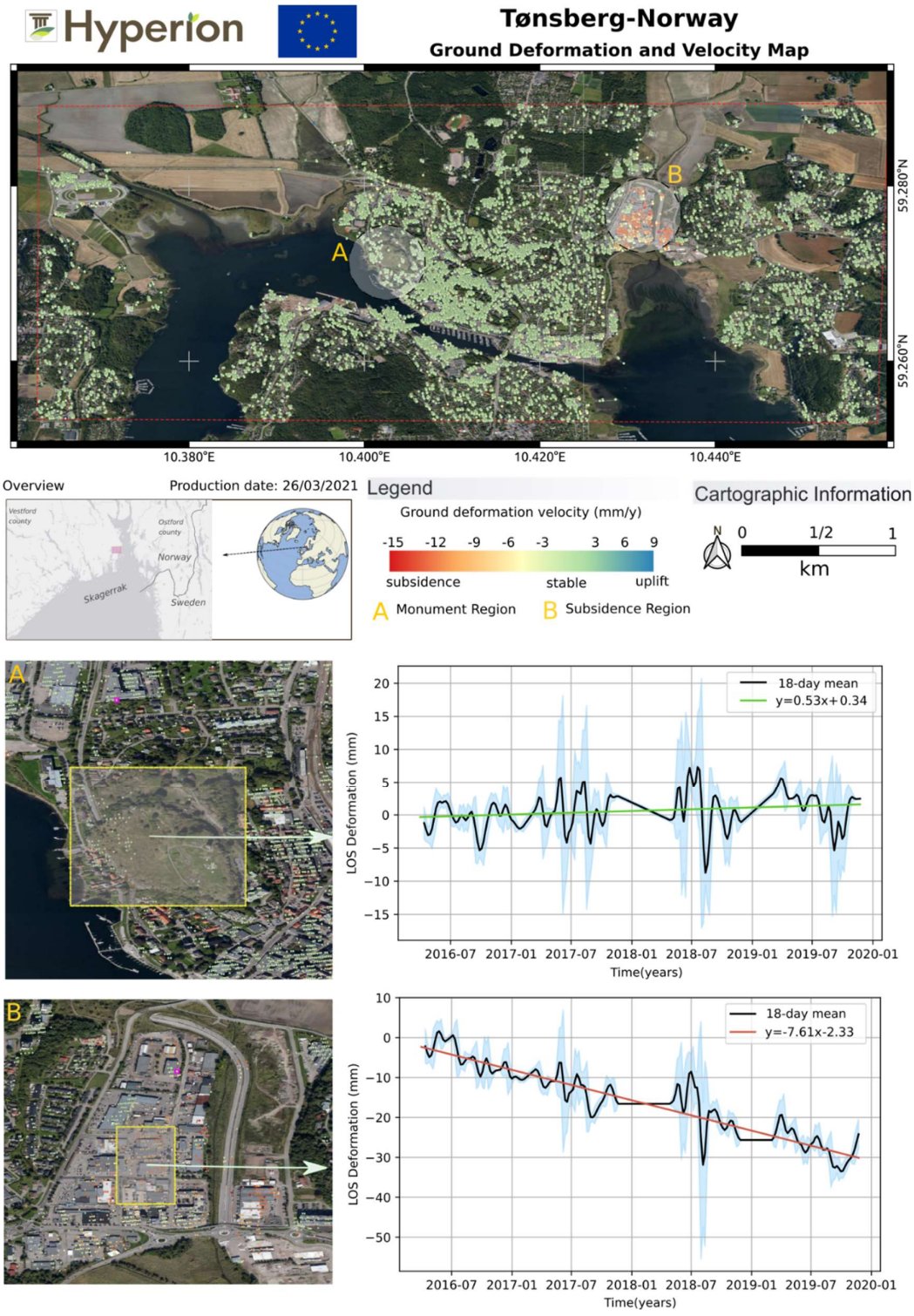

Figure 2: Ground deformation and velocity map for Tonsberg pilot area.

The ground deformation rate map using radar satellite data over an extended area of the Tonsberg pilot area is presented. The adopted technique was developed based on ISCE (Rosen et al., 2012) and STAMPS-PS (Hooper et al. 2004) functionalities for the detection of millimetre accuracy surface displacements. The used datasets consist of interferometric products based on Sentinel-1 dataset of 84 images (ascending stack spanning from March 2016 to December 2019) and SRTM 1 arc-second DEM data. Upon estimating the diachronic ground deformation pattern for a period from March 2016 to December 2019, district locations prone to subsidence are identified (area B).

Deformation results are referring to the Line of Sight (LOS) sensor to target direction. As a reference deformation value (zero value), the mean deformation of the AOI is considered. The deformation history for selected scatterers that are close (gray circles) to the CH monuments and present subsidence is plotted (Figure 2). The geolocation accuracy of each scatterer varies with respect to the DEM error.

2.1.2 Granada

The ground deformation rate map using radar satellite data over an extended area of the Granada pilot area. The adopted technique was developed based on ISCE (Rosen et al., 2012) and Mintpy (Yunjun et al., 2019) functionalities for the detection of millimetre accuracy surface displacements. The used datasets consist of interferometric products based on Sentinel-1 dataset of 116 images (descending stack spanning from January 2016 to December 2019) and SRTM 1 arc-second DEM data. Upon estimating the diachronic ground deformation pattern for a period from January 2016 to December 2019, district locations prone to subsidence are identified.

Deformation results are referring to the Line of Sight (LOS) sensor to target direction. As a reference deformation value (zero value), the mean deformation of the AOI is considered. The deformation history for selected scatterers that are close (gray circles) to the CH monuments and present subsidence is plotted in the following Figure. The geolocation accuracy of each scatterer varies with respect to the DEM error.

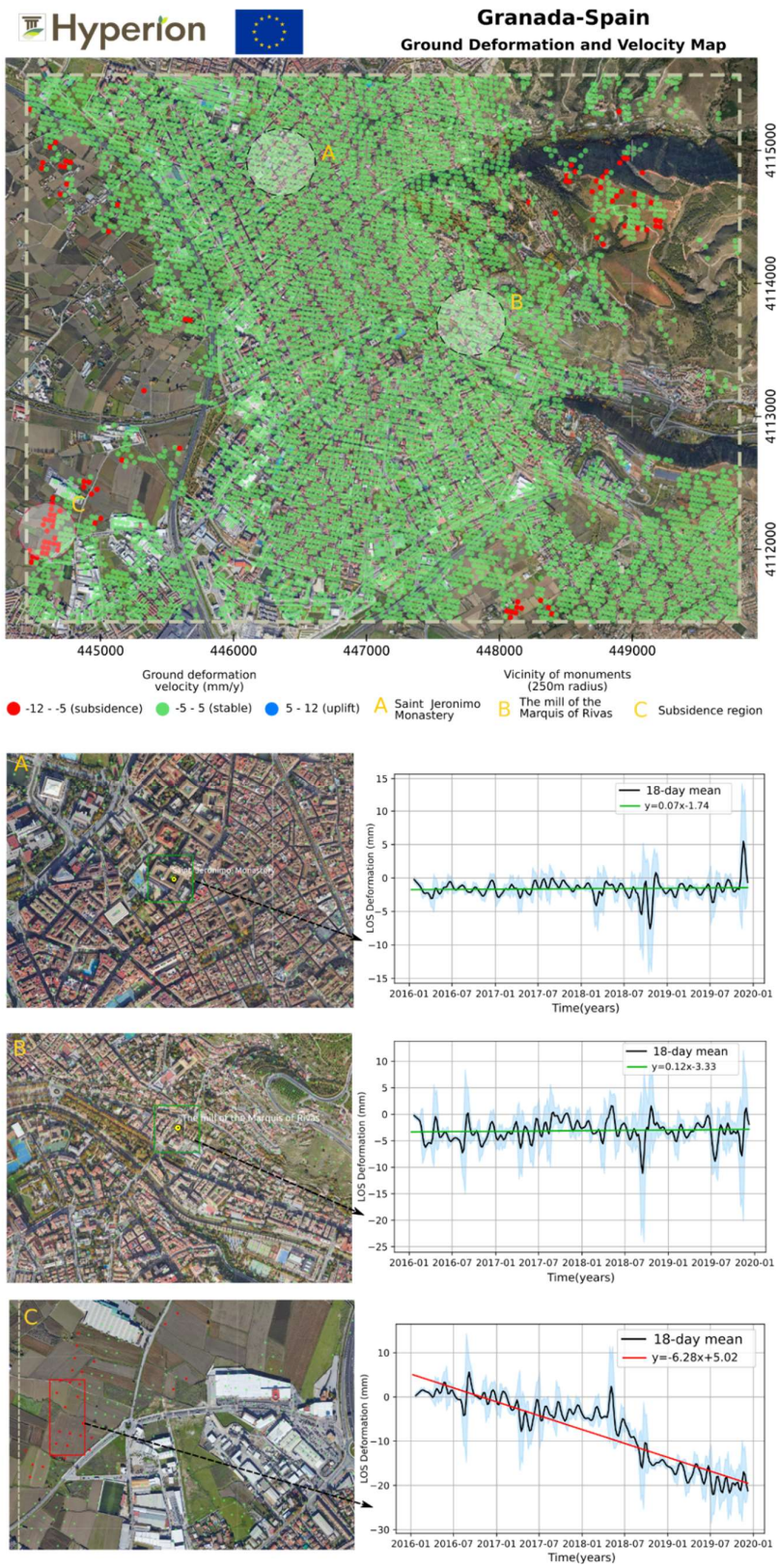

Figure 3: Ground deformation and velocity map for Granada pilot area.

2.1.3 Venice

The ground deformation rate map over an extended area using radar satellite data of the Venice pilot area. The adopted technique was developed based on ISCE (Rosen et al., 2012) and STAMPS-PS (Hooper et al. 2004) functionalities for the detection of millimetre accuracy surface displacements. The used datasets consist of interferometric products based on Sentinel-1 dataset of 106 images (descending stack spanning from February 2016 to December 2019) and SRTM 1 arc-second DEM data. Upon estimating the diachronic ground deformation pattern for a period from February 2016 to December 2019, district locations prone to subsidence and uplift are identified.

Deformation results are referring to the Line of Sight (LOS) sensor to target direction. As a reference deformation value (zero value), the mean deformation of the AOI is considered. The deformation history for selected scatterers that are close (gray circles) to the CH monuments with existing subsidence and uplift are plotted (Figure 4). The geolocation accuracy of each scatterer varies with respect to the DEM error.

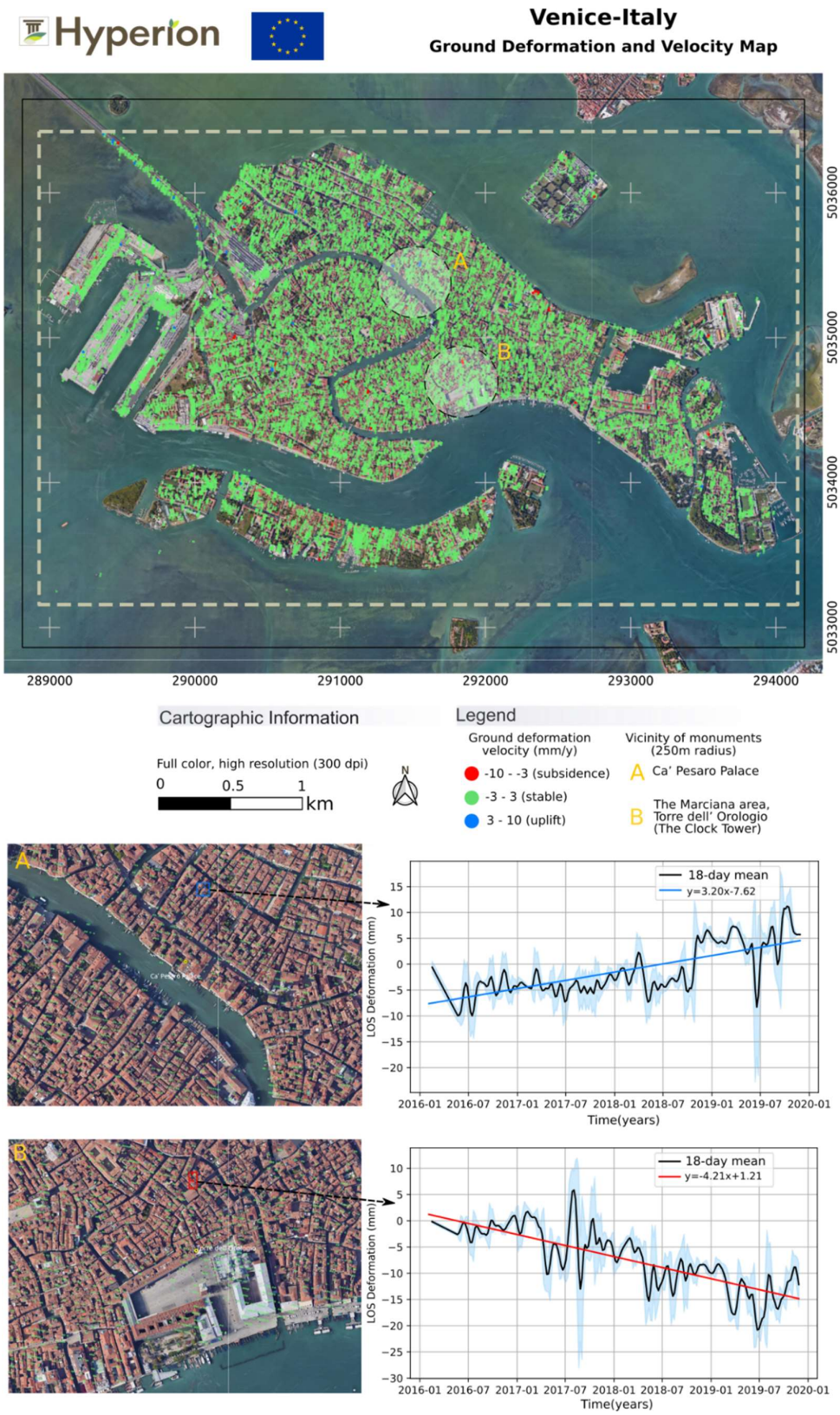

Figure 4: Ground deformation and velocity map for Venice pilot area.

2.1.4 Rhodes

The ground deformation rate map using radar satellite data over an extended area of the Rhodes pilot area. The adopted technique was developed based on ISCE (Rosen et al., 2012) and STAMPS-PS (Hooper et al. 2004) functionalities for the detection of millimetre accuracy surface displacements. The used datasets consist of interferometric products based on Sentinel-1 dataset of 106 images (descending stack spanning from January 2016 to November 2019) and SRTM 1 arc-second DEM data. Upon estimating the diachronic ground deformation pattern for a period from January 2016 to November 2019, district locations prone to uplift are identified.

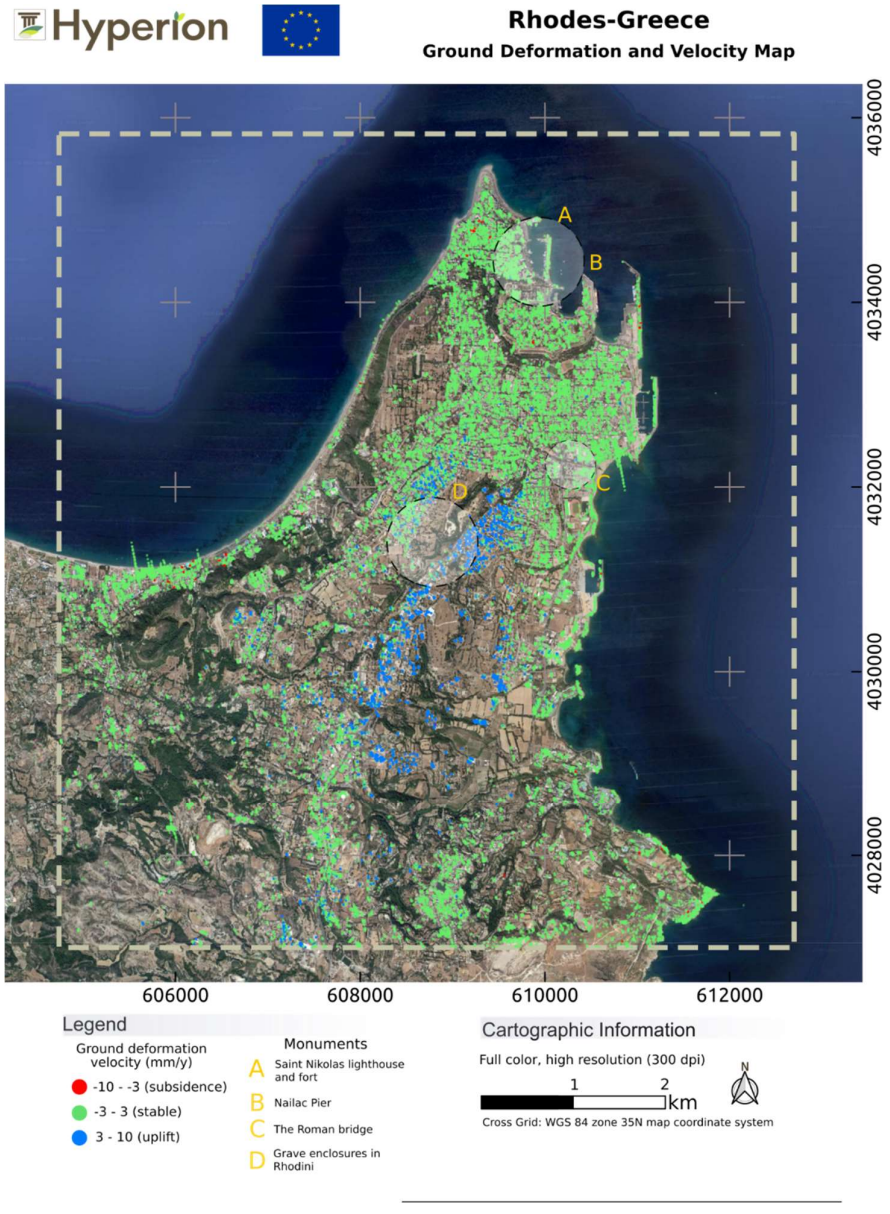

Figure 5: Ground velocity map for Rhodes pilot area.

Deformation results are referring to the Line of Sight (LOS) sensor to target direction. As a reference deformation value (zero value), the mean deformation of the AOI is considered. The deformation history for selected scatterers that are close (gray circles) to the CH monuments are plotted (Figures 5 and 6). The geolocation accuracy of each scatterer varies with respect to the DEM error.

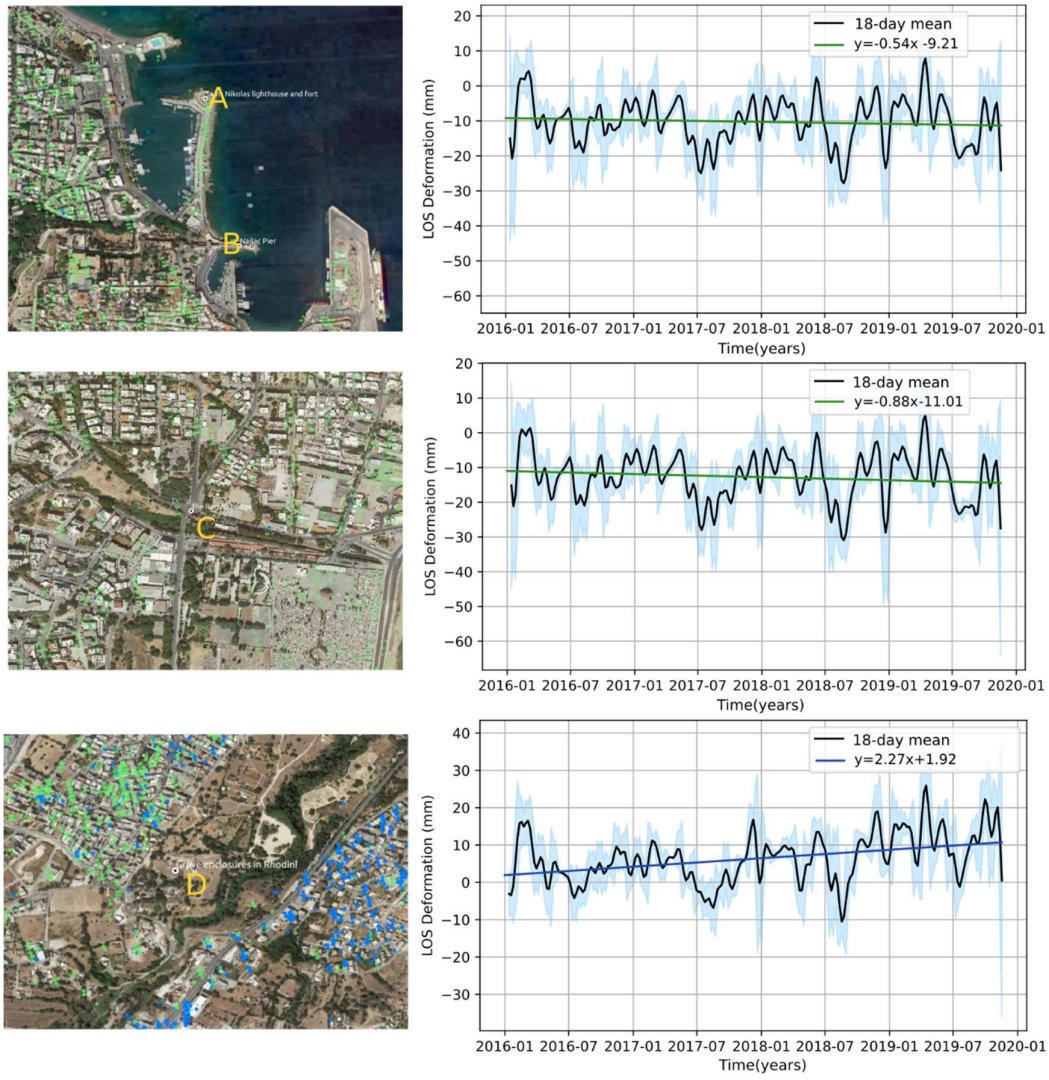

Figure 6: Ground deformation maps for Rhodes pilot area.

2.2 Land cover changes

2.2.1 Detailed information of procured images

For the detection of the land cover changes, very high resolution (VHR) pan-sharpened images collected from Geoeye-1 (GE01) and Worldview-2/3 (WV-2/3) satellites were used. The images were globally co-registered and contained spectral information in the visual and near-infrared (VNIR) part of the light spectrum. Their time difference varied between five and six years and the area size between 17 and 33 km^2. The spatial resolution for GE01 and WV-2 images was 0.5 m, while for WV-3 was 0.3 m. Details about the images are shown in Table 1 .

Table 1: Detailed information of VHR satellite images used for the land cover change detection

Area	Collection date	Satellite	x-size (px)	y-size (px)	Resolution (m)	Area (km^2)
Tonsberg	20/9/2013	WV-2	11873	8258	0.5	25
	12/7/2019	GE01	11873	8258	0.5	
Granada	19/7/2013	GE01	8094	10335	0.5	21
	2/7/2018	WV-3	13491	17226	0.3	
Rhodes	23/7/2013	WV-2	11337	11478	0.5	33
	5/6/2019	WV-3	18896	19131	0.3	
Venice	4/5/2013	GE01	10640	6533	0.5	17
	13/5/2018	WV-2	10640	6533	0.5	

2.2.2 Pre-processing steps

Before implementing the change detection methodology, the pre-processing steps were applied. These steps included: a) creation of mosaics from the WV-3 images since the area of interest was depicted in more than one tiles, b) resampling of the WV-3 images from 0.3 m to 0.5 m spatial resolution (same as GE01, WV2), and c) co-registration.

Four methods were tested for the automatic co-registration of the images. These methods were SIFT (Scale Invariant Feature Transform) (Lowe, 2004), ORB (Oriented FAST and Rotated BRIEF) (Rublee et al. 2011), a CNN feature-based approach (Yang et al., 2018), and the Fourier-Mellin Transform (Guo et al., 2005).

It is important to note that the procured images were not orthorectified, thus the co-registration process was applied locally and not globally. The local approach is necessary because of the perspective view geometry that causes non-uniform scale according to the relief. In more detail, SIFT, ORB, and the Fourier-Mellin transformation were tested on samples of size 1120 x 1120 px, while the CNN feature-

based approach was tested on patches of size 224 x 224 px. It should be also noted that because of the different satellite view angles, the images cannot be co-registered with high accuracy (e.g. visible/non-visible facades) (**Error! Reference source not found.**7).

Figure 7: Example of visible/non-visible facades in Venice because of the different satellite view angles. (a) Image collected on 13/5/2018 by WV-2. (b) Image collected on 4/5/2013 by GE01

SIFT locates local features known as “keypoints” that are scale and rotation invariant. The keypoints are detected by creating different scales of the images (application of Gaussian blur) and locating local maxima and minima. Then, their orientation and magnitude are defined by calculating gradients. Thus, a unique fingerprint is created for each point called “descriptor”.

ORB is a fusion of FAST (Features from accelerated segment test) (Rosten et al. 2005) keypoint detector and BRIEF (Binary Robust Independent Elementary Features) (Calonders et al. 2010) descriptor with modifications to enhance the performance. FAST is a corner detection method and BRIEF assigns descriptors by selecting a random pair of pixels in the neighborhood of a keypoint from a Gaussian distribution and comparing their brightness. The FAST modifications refer to the use of a multi-scale image pyramid and the assignment of orientation, while the BRIEF modifications to the inclusion of orientation invariance.

For both SIFT and ORB, the descriptor of one feature in the first set is matched with all other features in the second set using some distance calculation. During the matching process, outliers are excluded by the RANSAC (Random Sample Consensus) (Fischler & Bolles, 1981). In our case, for both methods, many points were incorrectly matched. An example area showing incorrectly matched points detected by the SIFT method is shown in Figure 8.

The CNN feature-based approach uses a CNN to generate multi-scale feature descriptors and then the Expectation Maximization method (EM) (Dempster et al., 1977) is applied to gradually increase the selection of inliers. This method produced distorted outputs especially in the edges of the patches (Figure 9).

Figure 8: Example area in Venice showing incorrectly matched points detected by SIFT. (a) Image collected on 13/5/2018 by WV-2. (b) Image collected on 4/5/2013 by GE01

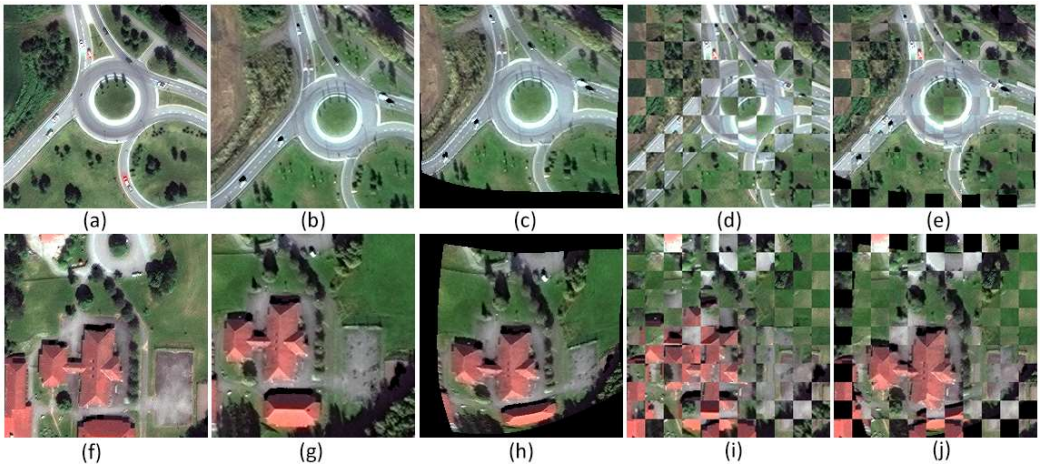

Figure 9: Example outputs of the CNN feature-based co-registration (Tonberg). (a, f) Image collected on 12/7/2019 by GE01. (b, g) Image collected on 20/9/2013 by WV-2. (c, h) Co-registered output. (d) Checkboard display of a & b. (e) Checkboard display of a & c. (i) Checkboard display of f & g. (j) Checkboard display of f & h.

In Fourier-Mellin Transform, the Fast Fourier transformation (FFT) of the input images is calculated and the magnitudes are mapped to log-polar coordinates. FFT is applied to the log-polar coordinates and then the rotation, scale-factor, and translation are retrieved by phase correlation. This method performed better than the others, but still not as well as the manual approach where matching points are manually collected. Figure 10 shows a comparison of co-registered outputs produced by the Fourier-Mellin Transform and the manual approach. It is shown that the Fourier-Melin Transform shows lower accuracy in areas of variable relief.

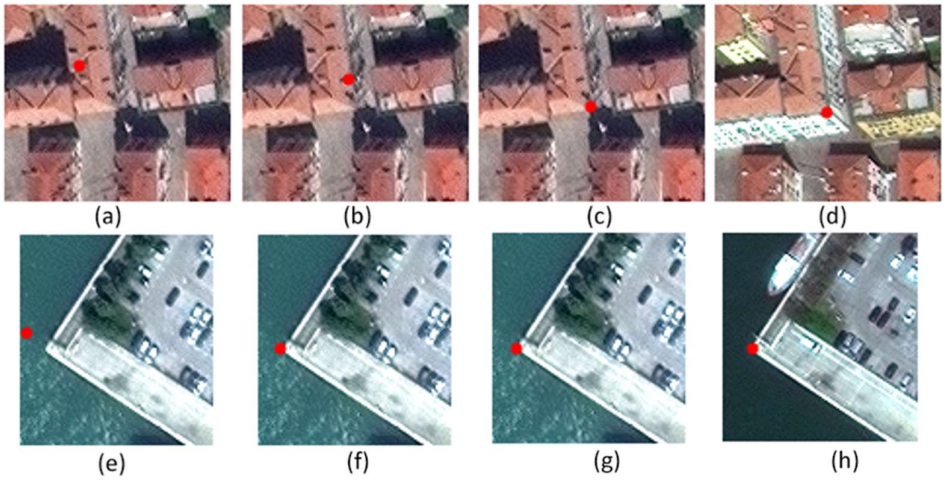

Figure 10: Comparison of Fourier-Mellin Transform and manual co-registration (Example outputs in Venice). (a, e) Image collected on 13/5/2018 by WV-2. (b, f) Co-registered output of Fourier-Mellin Transform. (c, g) Manually co-registered output. (d, h) Image collected on 4/5/2013 by GE01. The red bullet shows the position for a point.

Taking into consideration the above conclusions, it was decided to co-register the images manually. A grid with cells of size 1120 x 1120 px was created for each image and matching points were selected manually for 261 grid cells in total (Granada: 70, Rhodes: 59, Venice: 48, Tønsberg: 84). At least four points were selected for each grid cell and then the affine transformation was applied. Figure 11 shows the box plots of the Root Mean Squared Error (RMSE) for the four areas of interest. It can be seen that Granada showed the highest co-registration errors followed by Tønsberg, Venice, and Rhodes.

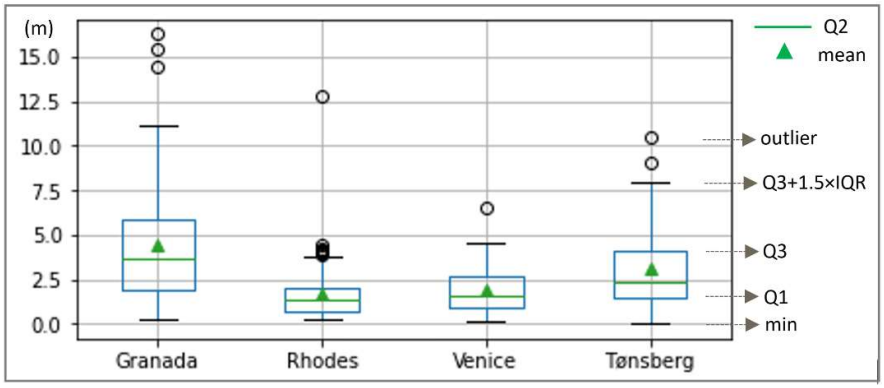

Figure 11: Box plots showing the distribution of the RMSE for the four areas of interest.

2.2.3 Unsupervised land cover change detection

Two unsupervised land cover change detection methods were implemented.

The first one applied the method proposed by El Amin et al. (2016) in which the feature maps produced by images collected in different dates are subtracted and after applying a threshold the change map is created. The proposed CNN uses the VGG-19 architecture (Simonyan & Zisserman, 2014) and is pre-trained on the Imagenet database (Deng et al, 2009) which contains millions of annotated RGB images with

thousands of categories. The size of the input image patches is 224 x 224 px and the output size is 112 x 112 px. The feature maps are extracted from five convolutional layers. The network was implemented by use of the Keras library (Chollet, 2015). The method is shown in Figure 12 and the result for an example area in Tønsberg is shown in Figure 13. The square shows the significant changes for our case. The results of this method show a high commission error caused by shadows, different satellite view angles (visible/non-visible facades), spectral dissimilarity of the same objects, and insufficient co-registration.

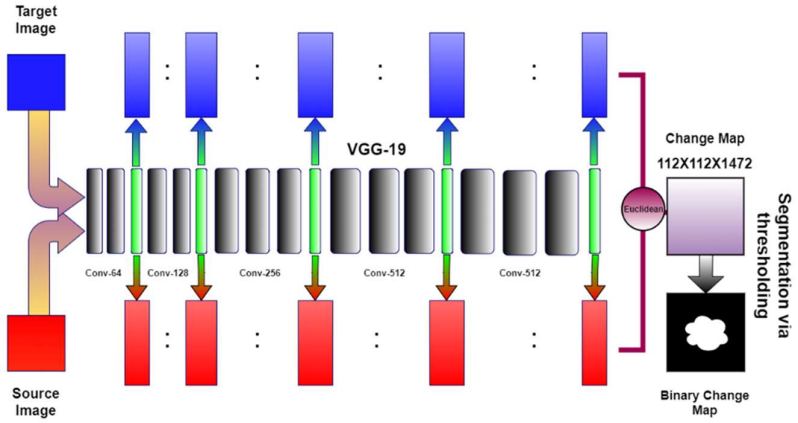

Figure 12: Unsupervised change detection approach proposed by El Amin et al. (2016)

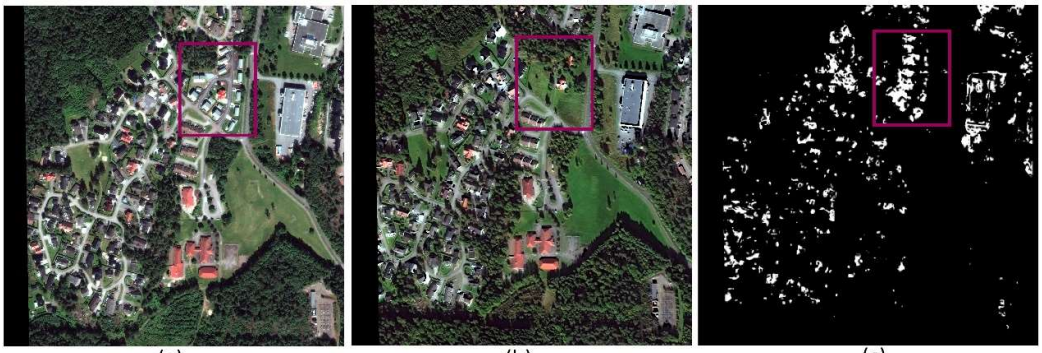

Figure 13: Example area in Tønsberg showing results produced by the method of El Amin et al. (2016). (a) Image collected on 12/7/2019 by GE01. (b) Image collected on 20/9/2013 by WV-2. (c) Result after thresholding. The square shows the significant changes.

The second unsupervised method applied a convolutional autoencoder with three convolutional layers in the encoder part (256, 128, 64 feature maps) and three convolutional layers in the decoder part (64, 128, 256 feature maps). The network was trained on patches (size:128 x 128) of the image of the first date and the visible and NIR bands were used. Then, similarly to the first unsupervised method, the feature maps were extracted and subtracted for both dates. The network was implemented by use of the Keras library. The result for an example area in Tønsberg is shown in Figure 14. As in the method proposed by El Amin et al. (2016), there is a high commission error caused by the same issues.

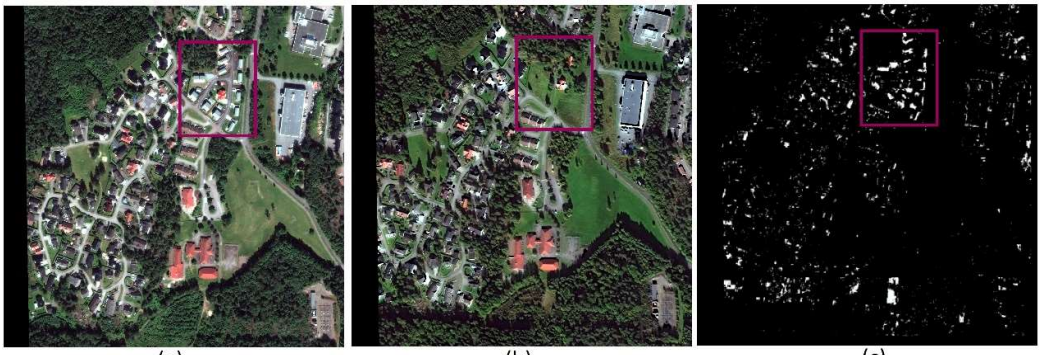

Figure 14: Example area in Tønsberg showing results produced by the autoencoder. (a) Image collected on 12/7/2019 by GE01. (b) Image collected on 20/9/2013 by WV-2. (c) Result after thresholding. The square shows the significant changes.

2.2.4 Supervised land cover change detection

Three supervised land cover change detection methods were implemented.

The first supervised method implemented the approach proposed by Zhang & Shi (2020) which uses transfer learning on a CNN (VGG-16) pre-trained on the AID dataset (Xia et al., 2017) (30 aerial scene types). The network (Feature difference CNN - FDCNN) contains two channels with shared weights. During the transfer learning, a small number of samples from VHR satellite images is used. The network was implemented by use of the Caffe framework (Jia et al., 2014). The flowchart of the approach is shown in Figure 15 and the result for an example area in Tønsberg is shown in Figure 16. The method shows a large commission error and it detects insignificant changes even in vegetation scenes.

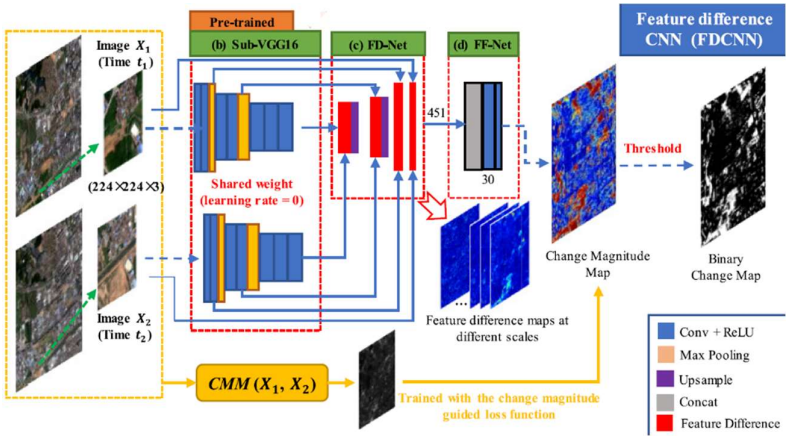

Figure 15: Flowchart of FDCNN (Zhang & Shi., 2020).

The second supervised method implemented the approach proposed by Chen et al. (2020) which uses dual attentive fully Convolutional Siamese networks (DASNet). The network was trained by its creators on a change detection dataset composed of multisource remote sensing images (spatial resolution of 3 to 100 cm/px) (10,000 training samples) (Lebedev et al., 2018) and it was implemented by use of the Pytorch

library.

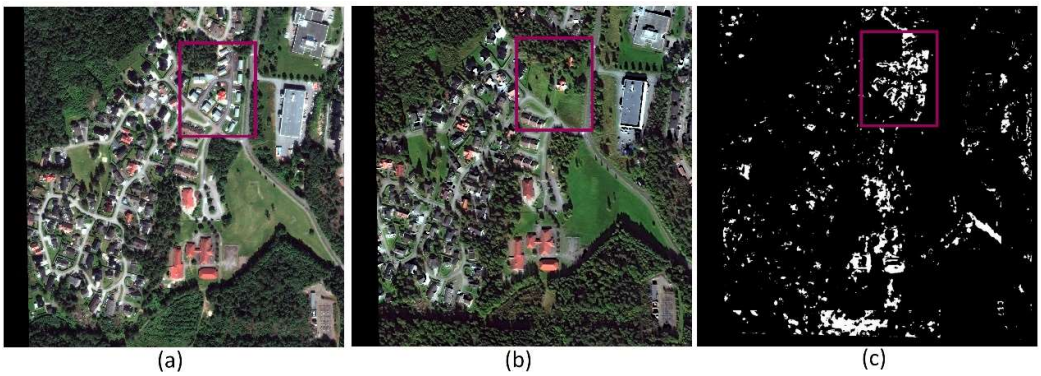

Figure 16: Example area in Tønsberg showing results produced by FDCNN. (a) Image collected on 12/7/2019 by GE01. (b) Image collected on 20/9/2013 by WV-2. (c) Result after thresholding. The square shows the significant changes.

An overview of DASNet is shown in Figure 17 and the result for an example area in Tønsberg is shown in Figure 18. The results show a large commission error and high sensitivity for spectral dissimilarity.

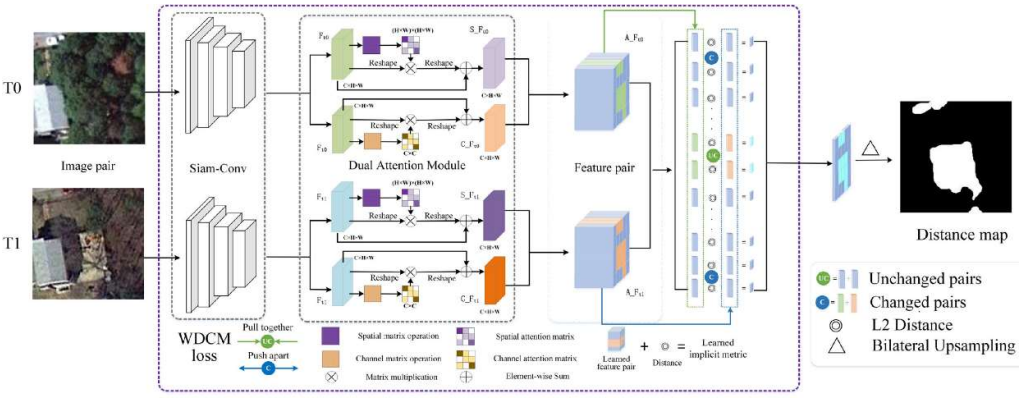

Figure 17: Overview of DASNet (Chen et al., 2020)

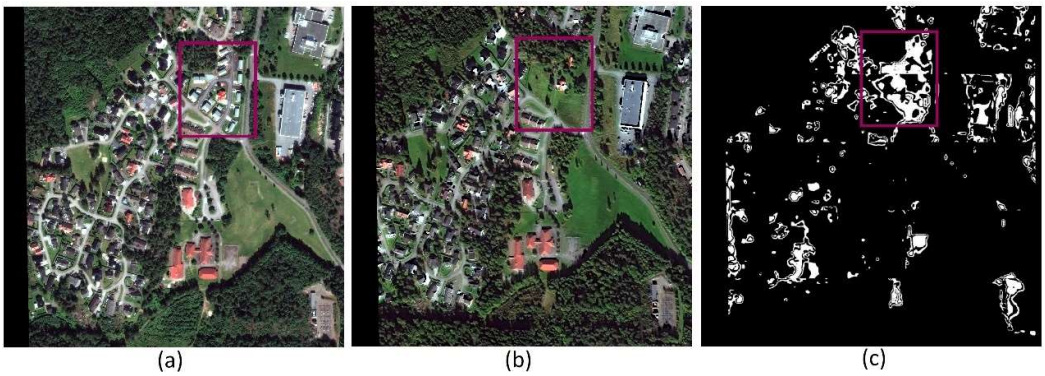

Figure 18 : Example area in Tønsberg showing results produced by DASNet. (a) Image collected on 12/7/2019 by GE01. (b) Image collected on 20/9/2013 by WV-2. (c) Result after thresholding. The square shows the significant changes.

For the third supervised method, the spatial-temporal attention-based network (STANet) proposed by Chen & Shi (2020) was applied. This network was trained on the LEVIR-CD dataset (proposed by the same researchers) which contains professionally annotated changes related to buildings (soil/grass/hardened ground → building). It was created from 637 VHR Google Earth image pairs (size: 1024 x 1024 px) from Texas, US and represents various types of buildings. Chen & Shi argue that their spatial-temporal mechanism creates more successful outputs in case of insufficient co-registration.

The pipeline of STANet is shown in Figure 19, Figure 20Figure 21, Figure 22, Figure 23 show the results for example areas in Tønsberg, Granada, Rhodes, and Venice respectively. It can be seen that changes related to buildings are detected more successfully than in all previously applied unsupervised and supervised methods. It can also be easily seen that the commission error is lower.

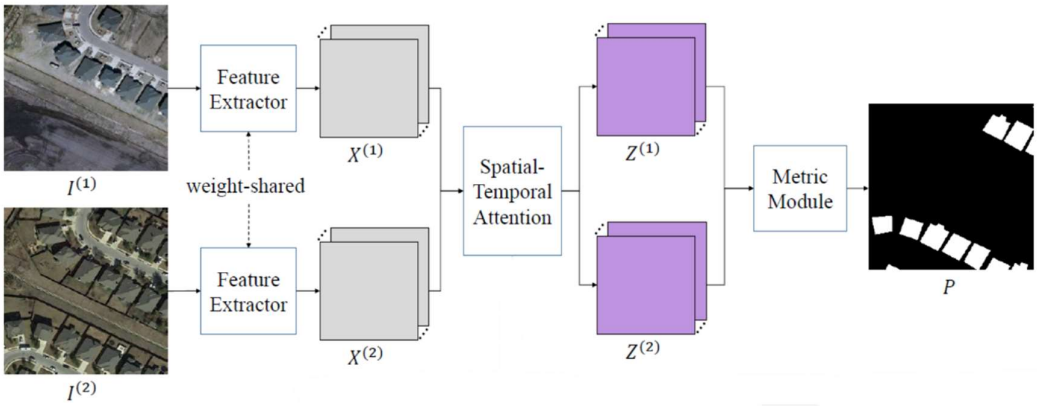

Figure 19: The pipeline of STANet (Chen & Shi, 2020)

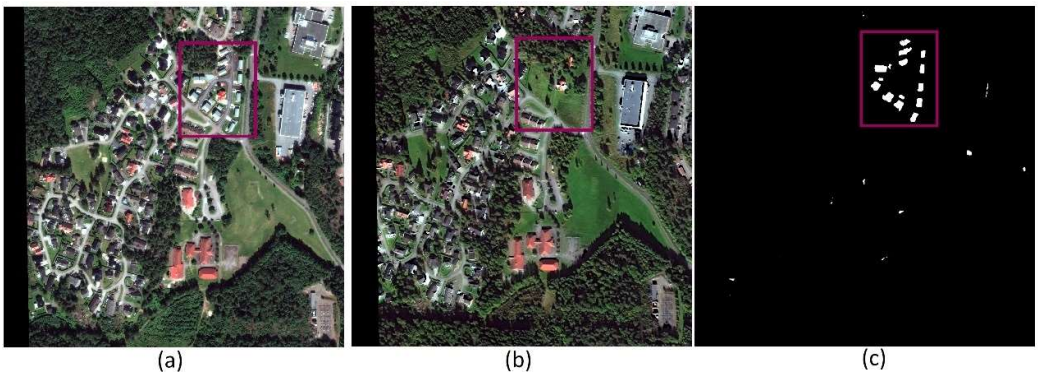

Figure 20: Example area in Tønsberg showing results produced by STANet. (a) Image collected on 12/7/2019 by GE01. (b) Image collected on 20/9/2013 by WV-2. (c) Result after thresholding. The square shows correct change instances.

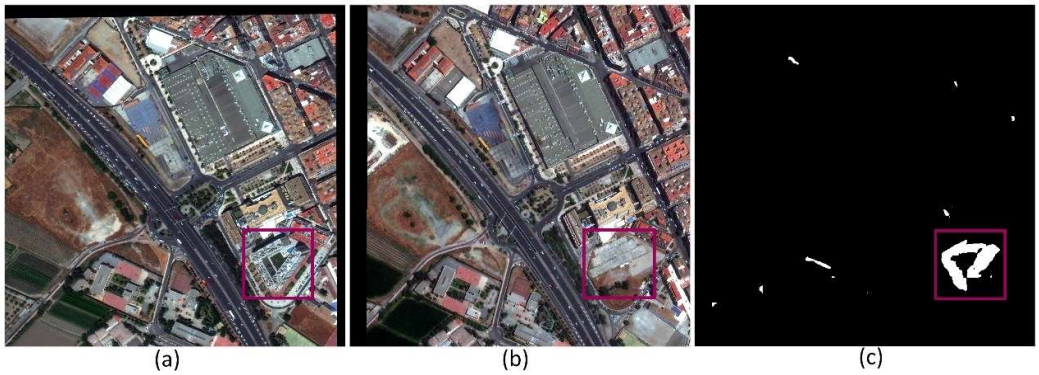

Figure 21: Example area in Granada showing results produced by STANet. (a) Image collected on 2/7/2018 by WV-3. (b) Image collected on 19/7/2013 by GE01. (c) Result. The square shows correct change instances.

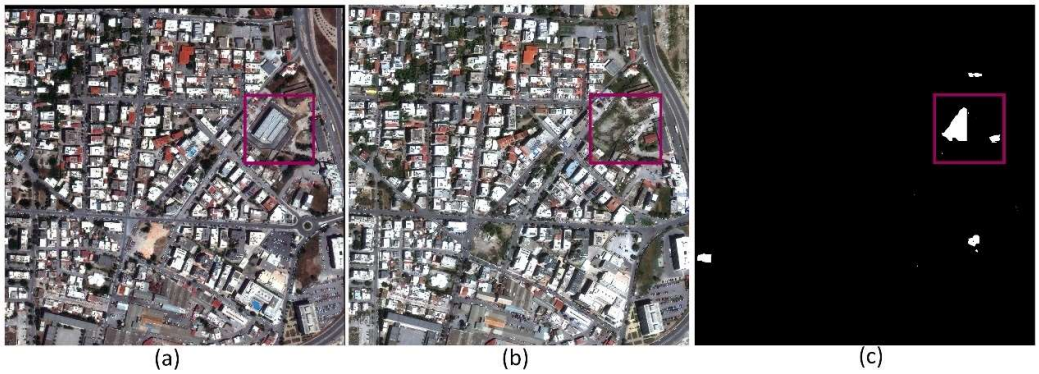

Figure 22: Example area in Rhodes showing results produced by STANet. (a) Image collected on 5/6/2019 by WV-3. (b) Image collected on 23/7/2013 by WV-2. (c) Result. The square shows the correct change instances.

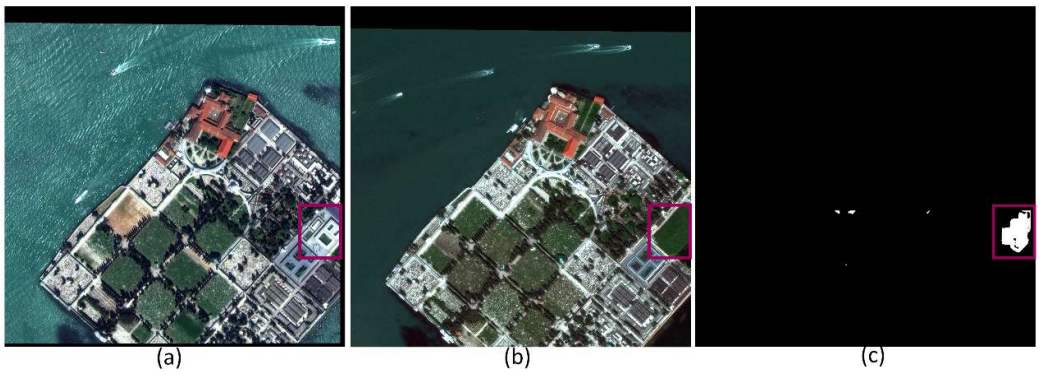

Figure 23: Example area in Venice showing results produced by STANet. (a) Image collected on 13/5/2018 by WV-2. (b) Image collected on 4/5/2013 by GE01. (c) Result. The square shows the correct change instances.

This method was implemented for the whole area of the four study areas and besides being evaluated qualitatively, it was also evaluated quantitively by the calculation of metrics. These metrics were Recall (True Positive/ (True Positive + False Negative)) which corresponds to omission error and Precision (True Positive/ (True Positive + False Positive)) which corresponds to commission error. The evaluation metrics for all

four study areas and the training set (reported by the creators of STANet) are shown in Table 2.

By observing the Table 2 it can be seen that the omission error is lower than the commission error. The lowest omission error is presented in Rhodes (6%) and the highest in Venice (20%). The commission error is higher than 30% for all study areas and was caused mainly from the co-registration errors and the different satellite view angles. This error percentage is expected since in much better conditions (training set composed of images from the same satellite with small co-registration errors) the network showed 16% commission error.

Table 2: Evaluation metrics for the results of STANet

Area	Recall	Precision
Rhodes	94%	71%
Venice	80%	57%
Tønsberg	88%	66%
Granada	90%	51%
Training set	91%	84%

The pie charts displayed in Figure 24 show the percentages of the types of changes detected by STANet for the four study areas. The highest pseudochanges are presented in Granada and Venice because of the presence of high building blocks and the different view angles of GE01 and WV. Another challenge for Granada was its mountainous terrain. The lower pseudochanges for Rhodes can be attributed to the similar view angle of WV-2 and WV-3, while for Tønsberg to the low building height and higher similarity with the training set. It should be noted that as shown in the box plots of Figure 11, Granada presented the highest co-registration error, while Rhodes the lowest.

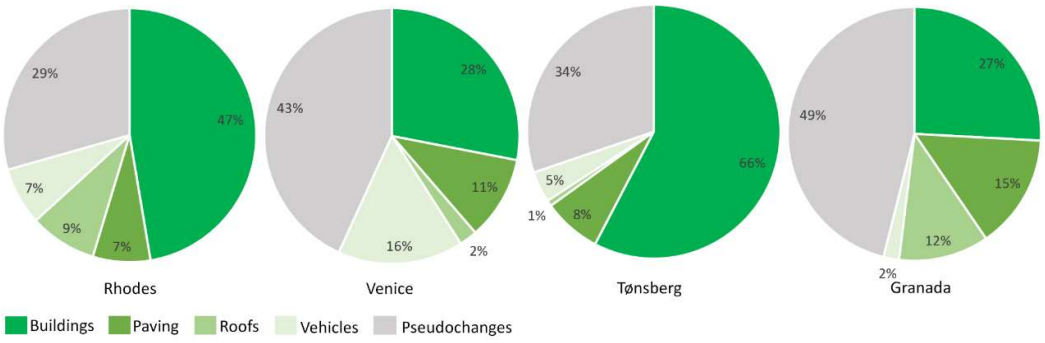

Figure 24: Percentages of the types of changes detected by STANet

Figures 25, 27, 29, 31 show the whole area covered by the satellite images for the four study areas. Figures 26, 28, 30, 32 show the change maps produced by the implementation of STANet.

Figure 25: Satellite image of the Tønsberg study area collected on 20/9/2013 by WV-2. Area size:25 km²

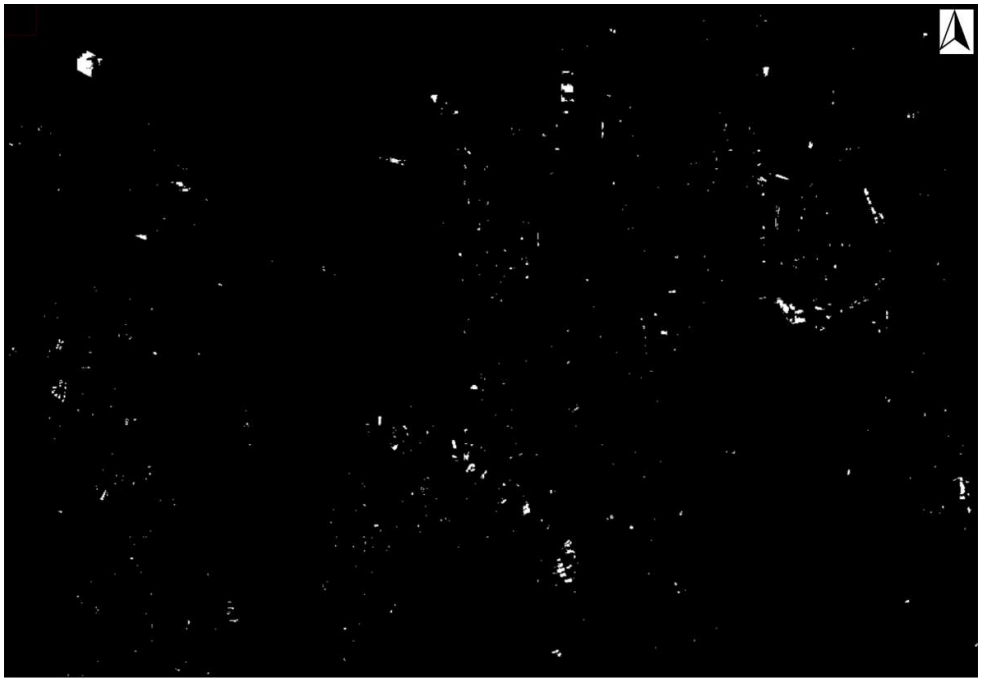

Figure 26: Change map of the whole Tønsberg study area produced by STANet.

Figure 27: Satellite image of the Granada study area collected on 19/7/2013 by GE01. Area size:17 km²

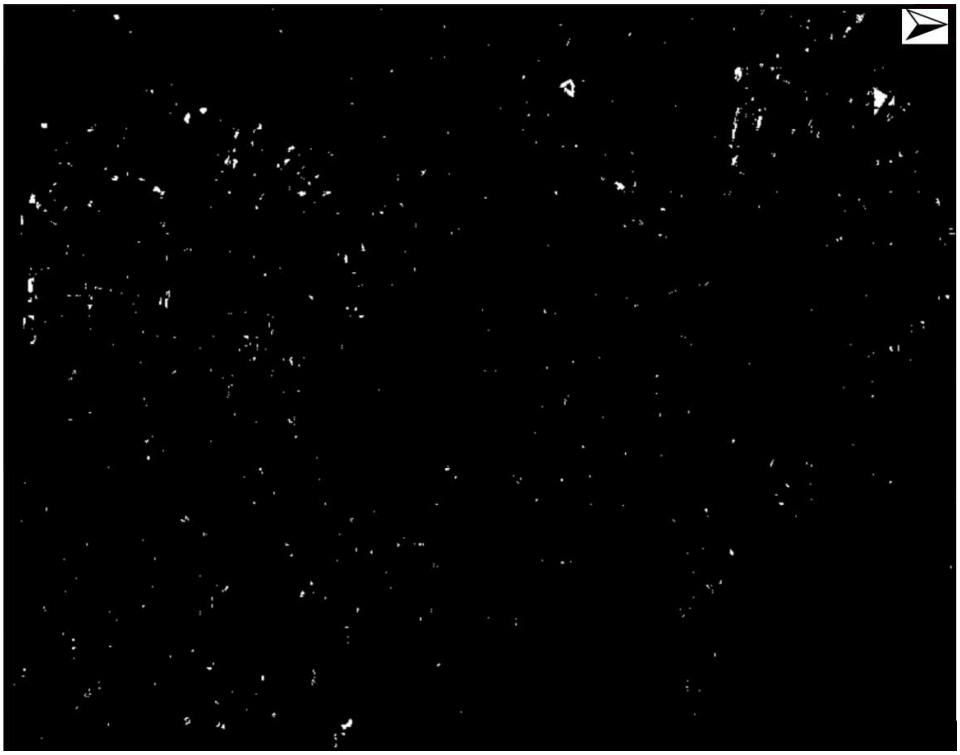

Figure 28: Change map of the whole Granada study area produced by STANet.

Figure 29: Satellite image of the Rhodes study area collected on 23/7/2013 by WV-2. Area size:33 km^2

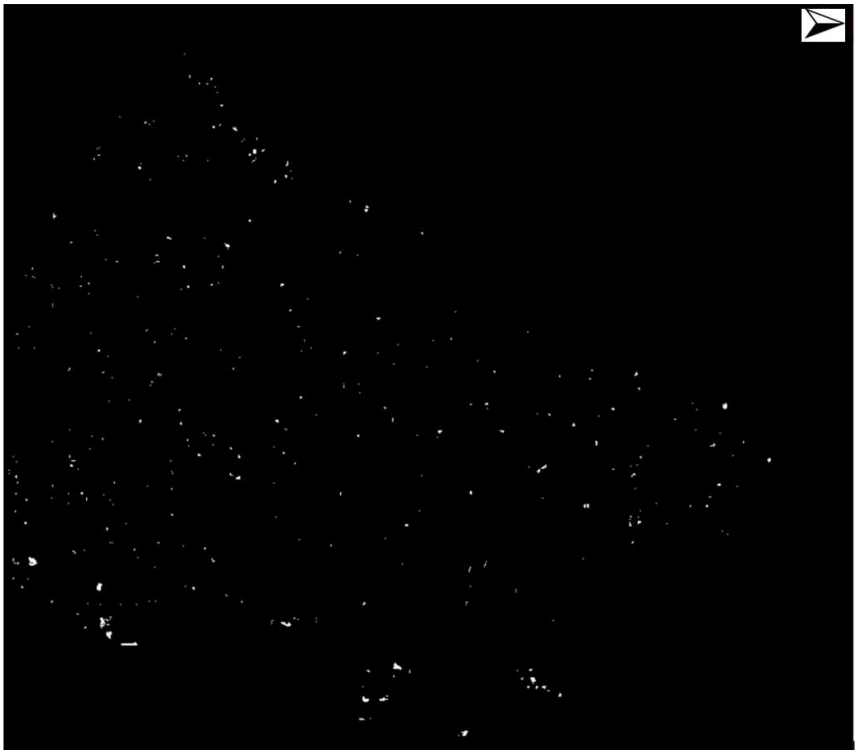

Figure 30: Change map of the whole Rhodes study area produced by STANet.

Figure 31: Satellite image of the Venice study area collected on 4/5/2013 by GE01. Area size:17 km²

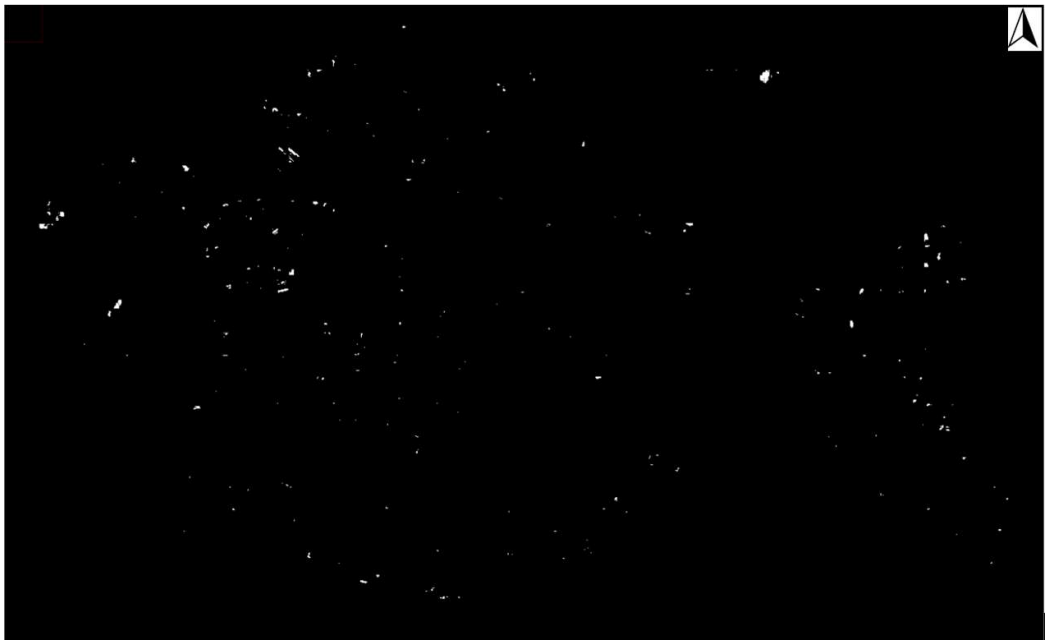

Figure 32: Change map of the whole Venice study area produced by STANet.

3. 3D-representation of the CH assets

For the initial 3D modelling and representation of the Cultural Heritage buildings the combination of geodetic, photogrammetric and laser scanning data acquisition and processing methods have been applied. Digital images have been taken in different ways according to the size, complexity and level of detail of each monument in Rhodes using high resolution full-frame cameras and two UAS and with two low resolution multispectral cameras (Tapinaki et al., 2021). Furthermore, terrestrial laser scanning has been conducted to accurately determine the surface of the Cultural Heritage buildings used as reference for the routine and post-disaster monitoring, as foreseen by the project. The objective is to produce at first dense point clouds for each monument and then detailed, textured 3D meshes, light 3D models and sections.

All the final products were uploaded at the project's sharing point so that all the partners could have access and make use of them for further analysis (https://isense-cloud.iccs.gr/s/a2TJoer9QbPnqnM/download). The point clouds, the detailed, textured 3D meshes and the light 3D models will also be a part of the HRAP Platform.

3.1 Point clouds

The data processing procedure focused on the exploitation of both the photogrammetric data and the laser scanning data in order to produce a complete and uniform point cloud. First the digital images were processed using the Image Based Modelling (IBM) software Agisoft Metashape v.1.6.5 following the standard workflow as in every documentation process. The captured images for each pilot area were loaded in chunks, separating the aerial and the terrestrial images. The images were examined, evaluated for their quality and usefulness and some of them were excluded. The next step was the automatic detection of the pre-marked targets and the manual detection of the laser scanner targets to facilitate the merging of the chunks, the alignment, scaling and georeferencing of each 3D point cloud. After the successful alignment of the images (Figure 33), the dense point cloud was generated and subsequently extracted to be further processed with the Geomagic Wrap 2017 software.

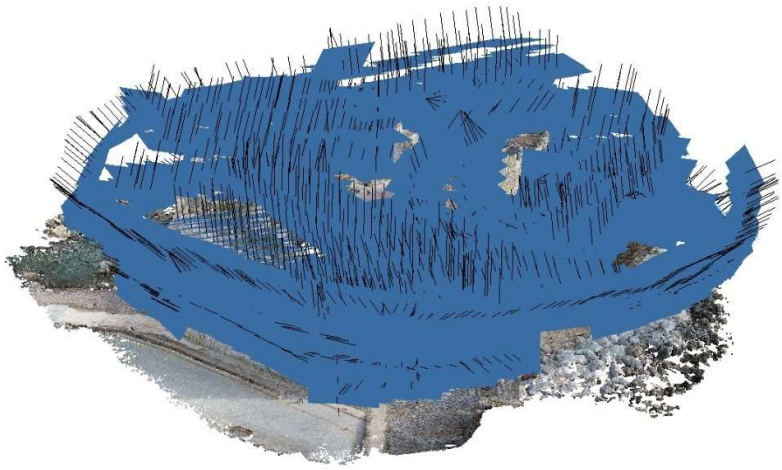

Figure 33: The result of the alignment process in the IBM software and the position of the cameras for the St. Nikolas Fort in Rhodes.

Then the data processing continued with the registration of the scanned point clouds using the Faro Scene and Cyclone Register 360 (BLK Edition) software applying the Cloud-to-Cloud method since the overlap between the scans was large enough, while the aligned scans were then georeferenced by identifying the targets and setting the coordinates. The point cloud of each Cultural Heritage building was further processed using the Geomagic Wrap 2017 in order to reduce inevitable scanner errors (noise) and reduce the number of points leading to a smoother and accurate 3D Model. Finally, the dense point cloud from the IBM software was used to fill the eventual gaps in the scans and generate the final point cloud. The dense point cloud from the IBM required more editing since it presents more noise and inaccurate points. Various filters were applied in order to create an accurate point cloud and after that it was combined at each case with the laser scanner point cloud to be converted to a 3D mesh (Figure 34). These point clouds will be used as reference for the routine and post-disaster monitoring after every 3D documentation to perform comparisons and evaluations of the deviation between these point clouds.

Figure 34: The final, combined dense point cloud for the Naillac Tower in Rhodes.

3.2 Sections

The detailed 3D Mesh for each Cultural Heritage building was further processed in Geomagic Wrap 2017 using the appropriate commands and filters to obtain the final 3D surface. This last process includes filling inevitable small holes, inversion of any inaccurate triangles and smoothing the sharp edges where it was necessary. The final detailed 3D surface was used to extract vertical and horizontal sections for each Cultural Heritage building with a 6cm interval. The sections were necessary for the project in order to develop the Finite Element Models and perform structural analysis. The final 3D surface was imported into the Rhino v.6 software and the vertical and horizontal sections were produced as contours with an interval of 6cm.

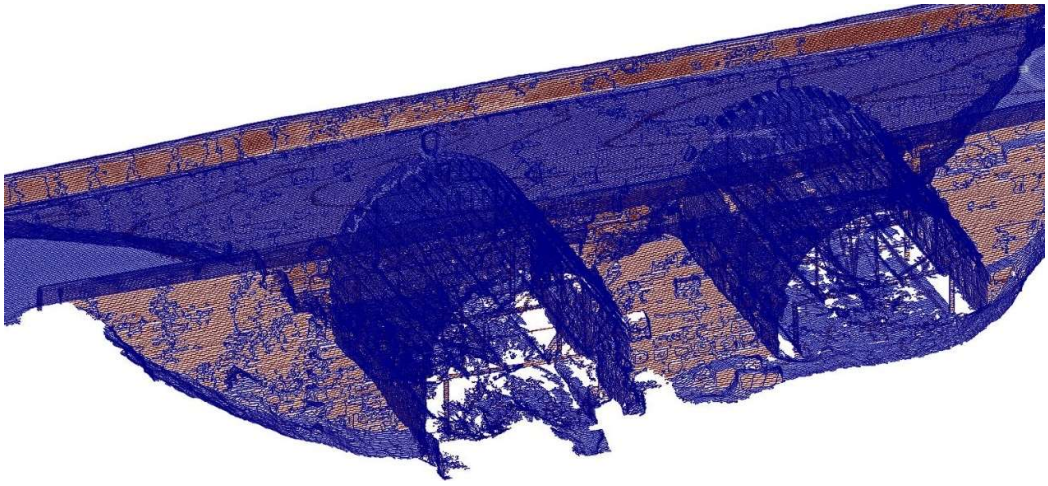

Figure 35: The vertical and horizontal sections for the Roman Bridge in Rhodes (6cm interval).

3.3 Detailed Textured models

The detailed high-resolution textured 3D models were produced for specific parts of each Cultural Heritage building in places where the hyperspectral data were also acquired. The scope was to integrate the RGB and HS data and get results for both the

geometry and the material loss for each case. It was decided to trim the final 3D surface and isolate the meshes that were covering the smaller areas of the HS data. The 3D meshes were imported into the IBM software to produce the 3D Textured model using only the necessary RGB high resolution digital images for each part. The texturing parameters used for this process were the generic mapping mode and the mosaic blending mode, while for each detailed textured model it was chosen to develop at least 4 texture maps (Figures 36 & 37). Moreover, an orthomosaic and the Digital Elevation Model (DEM) were produced for these parts.

Figure 36: The detailed 3D Textured model for the St. Nikolas Fort in Rhodes.

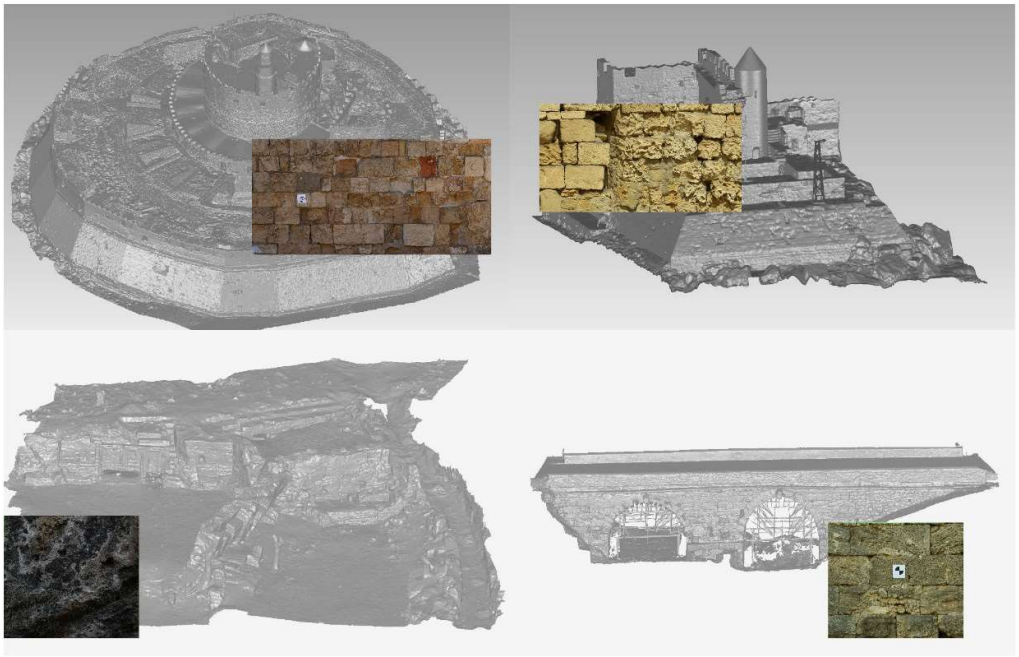

Figure 37: Some of the detailed 3D Textured models for the 4 pilot areas in Rhodes.

3.4 Light 3D Models

The final, detailed 3D surfaces of the Cultural Heritage buildings will be imported into to HRAP Platform for further analysis in combination with various data. For visualisation purposes these files had to be decimated to make the files lighter but without altering the geometry and keeping them accurate and detailed. For this reason, the 3D meshes were edited in the Geomagic software using the appropriate commands and reducing the triangles. Furthermore, the vertex colour of the meshes was removed (Figure 38).

Figure 38: The shaded 3D model (Left) and the 3D model without colour (Right) of the Naillac Pier in Rhodes.

4. Mapping the material deterioration of the CH assets

4.1 Spectral Library

The hyperspectral data acquisition has been carried out using the HyperView multi sensor hyperspectral sensing platform by 3D-one. The details of the hyperspectral system used and the necessary pre-processing have been documented in deliverable D6.2. The specific hyperspectral system, which was the only available to be used in the project, has limited spectral resolution. Nevertheless, interesting findings on material deterioration were derived from the hyperspectral images that have been acquired.

The pre-processed hyperspectral (HS) images have been used for the extraction of the spectral signature of the Cultural Heritage building materials and biodeterioration appearances.

Spectral signatures of various samples from the hyperspectral images have been studied. Signatures have been collected for (1) building material types, (2) bio-degradation types or other corrosion factors according to the ICOMOS-ISCS illustrated glossary on stone deterioration patterns (ICOMOS-ISCS, 2011), and (3) material loss condition. Small polygons have been selected for the calculation of the spectral signatures for pure appearances. Later on, these pure spectra were used as endmembers during spectral unmixing. The stone materials were classified according to their colour, based on the work of Papagianni et al, 2004. For the signatures of these materials differences were encountered mainly in the visual spectrum and more particular in the green-red region. The spectral signatures of the various corrosion factors presented more variations. Eflorescence, daub and salt crystallization presented a signature similar to this of the reference surface. Vegetation appearing on the walls presented a strong increase in the Near Infrared (NIR) spectrum with a slight drop around 880 nm. Moss, which was the most common bio-deterioration factor in the specific monuments, also presents increase in the NIR part of the spectrum but less than the leaves of the bigger plants. Dry moss presents only a slight increase in the NIR region.

As far as material loss is concerned, it was found that the spectral signatures present a general decrease in reflectance when the material loss is more intense (Figure 39). This reflectance decrease is caused by the rougher texture and the occurring shadows and not changes within the stone material. Thus, spectral signatures for material loss cannot be selected and will not be discussed further.

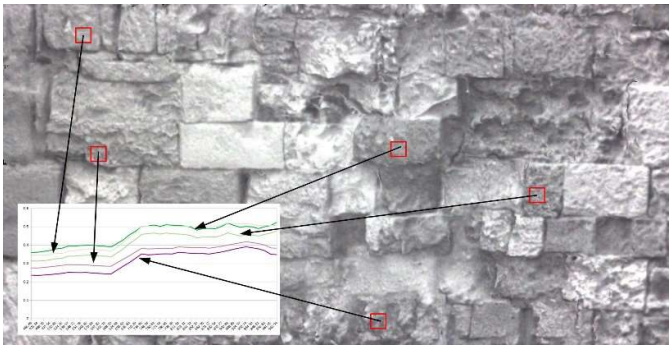

Figure 39: Detailed 3D Textured Model (HS) and spectral signatures of stones with various degradation levels.

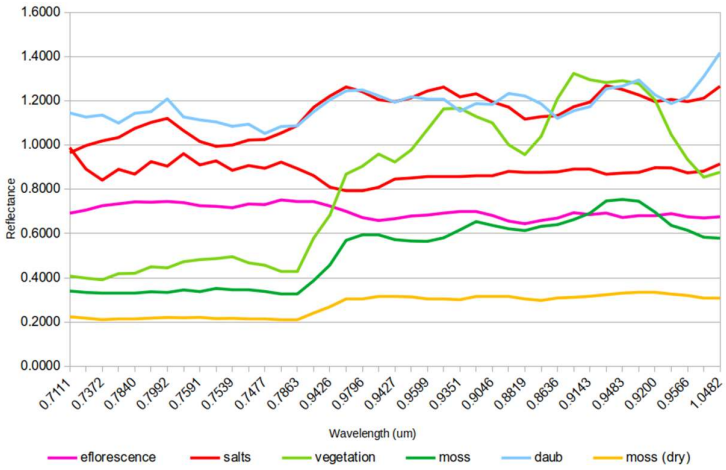

Figure 40: Spectral signatures derived from the HS images of the St. Nikolas Fort

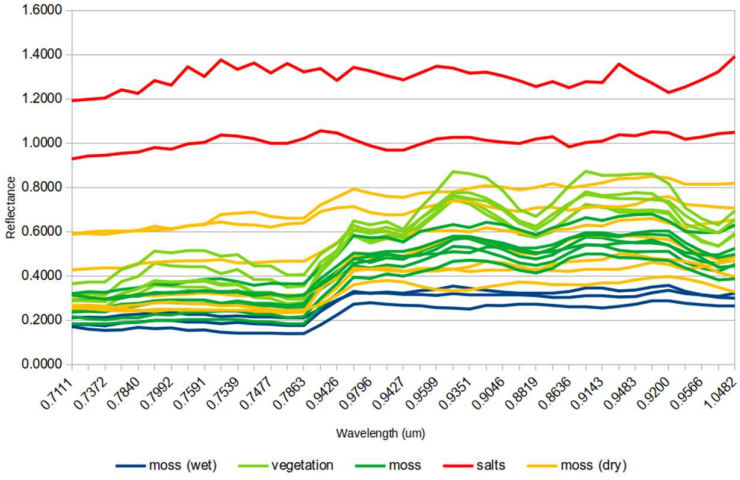

Figure 41: Spectral signatures derived from the HS images of the Roman Bridge

Figure 40 and Figure 41 present spectral signatures from the St. Nikolas Fort and the Roman Bridge respectively. In the case of the Roman bridge, some parts of the moss were wet, presenting very low reflectance especially in the NIR region. Signatures of

the same appearances present similar shapes but different reflectance magnitudes due to irradiance variations, which are still apparent despite the radiance to reflectance conversion.

A limited site-specific spectral library has been built using the extracted spectra for the previously mentioned corrosion factors as well as for the wall materials. This spectral library has been used for selecting appropriate endmembers during the supervised spectral unmixing process which has been carried out.

Continuum removed spectra has also been studied in order to detect useful absorption features, but it did not prove to be useful in this case study.

4.2 Deterioration mapping through spectral unmixing

The signal detected by a sensor into a single pixel is frequently a combination of numerous disparate signals that jointly occupy the pixel by forming a macroscopic mixture. These disparate signals correspond to distinct substances that are called endmembers or pure pixels and the fractions in which they appear in a mixed pixel are called abundances. A very common procedure when working with hyperspectral imagery is the unmixing process, which aims at detecting the endmembers in a given scene as well as their abundances. The accuracy of spectral unmixing is highly dependent on the selection of appropriate, representative endmembers (Tompkins 1997). In many cases, selection of appropriate endmembers is carried out manually. Then supervised unmixing process is performed.

The hyperspecral imagery collected in Rhodes has been processed with various unmixing methods (linear spectral unmixing, matched filtering, mixture tuned matched filtering, constrained energy minimization etc.) in a supervised way, providing the collected spectral signatures as endmembers. The unmixing method that provided the best results was the mixture tuned matched filtering (MTMF). MTMF is an advanced spectral unmixing algorithm in which it is unnecessary for all materials within a scene to be known or to have identified endmembers (Boardman et al. 1995), and it combines the best aspects of the linear spectral mixing model and the statistical matched filter model. MTMF performs partial unmixing only by finding the abundance of a single, user-defined endmember, by maximizing the response of the endmember of interest, and by minimizing the response of the composite unknown background, thus ‘matching’ the known signature (Williams and Hunt 2002).

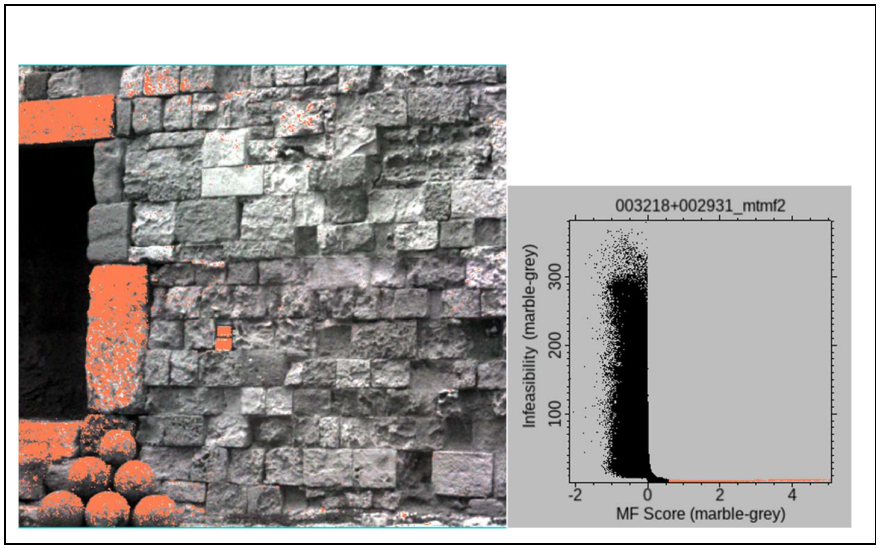

Figure 42: Paint in orange are the pixels which are selected for marble MF score > 0.5 and marble infeasibility < 20

The MTMF method includes three main steps (Mehra et al, 2013): (1) an MNF transformation of apparent reflection data (Green et al.1988), (2) matched filtering for abundance estimation, and (3) mixture tuning to identify infeasible or false-positive pixels (Boardman1998). The output of MTMF is a set of rule images given as MF and infeasibility scores for each pixel related to each endmember. In order to be sure that the abundance images contain only correct values, the pixels with high MF scores and low infeasibility values have to be selected. In our experiments, pixels with MF score above 0.5 and infeasibility below 20 have been selected as correct for each material type (Figure 42).

Spectral unmixing, MTMF in our case study, provides a set of abundance maps for all the endmembers (pure spectra) that have been used. By appropriate selection of thresholds someone can create a classification of the dominant endmember within every pixel. As already mentioned, in this work the MTMF unmixing results have been used for building material maps and biodeterioration maps. Thresholds of MF score > 0.5 and infeasibility < 20 have been applied on each abundance map for classification purposes. Such maps are shown in Figure 43.

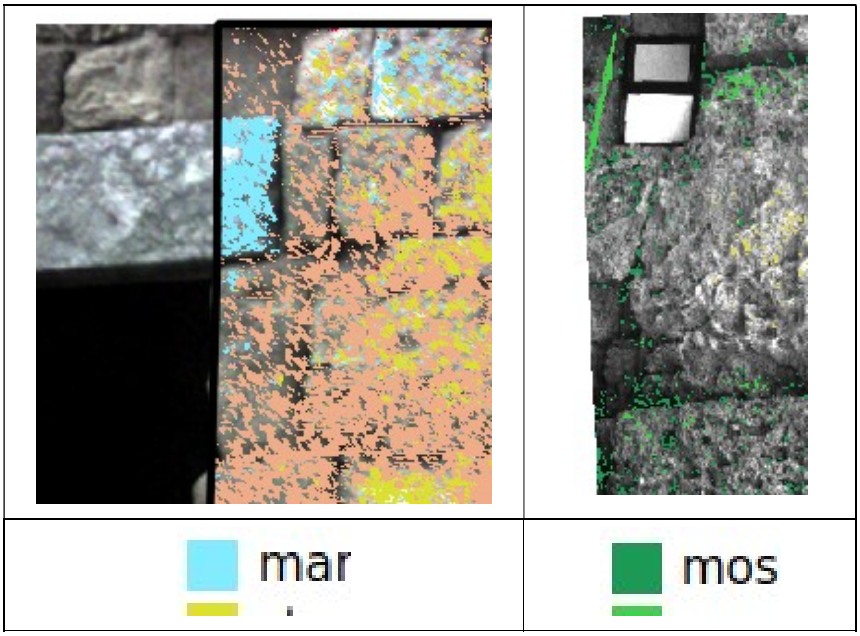

Figure 43: Left image: Building material map derived from MTMF spectral unmixing (St. Nikolas Fort). Right image: Biodeterioration map derived from MTMF spectral unmixing (Roman Bridge).

4.3 Material loss estimation

Both RGB and Hyperspectral data were acquired for specific parts of the 4 pilot areas in Rhodes. A first attempt was made in order to exploit and integrate both the close range 3D optical and the Hyperspectral data for the assessment of material degradation (Kolokoussis et al., 2021). For this reason, two approaches have been used, a planar-surface approach and a valley depth approach. For the first, more simple, approach the detailed textured 3D models for the smaller parts of the buildings were used as well as the dense point clouds. These data were imported into the CloudCompare Software (CloudCompare 2.12 alpha, 2021). The approach was based on the adjustment of a planar surface at the part of the monument that presented no degradation. After defining the plane, the distances between the plane and the 3D Point Cloud or 3D Mesh would indicate the depths of the degradation and the material loss for each stone at the Cultural Heritage buildings by using the qMplane Plugin. Numerous points were picked to accurately define the plane at the parts where no degradation was obvious according to the 3D Textured model (Figures 44 and 45).

Figure 44: The Detailed 3D Textured Model where the erosion is obvious at specific stones.

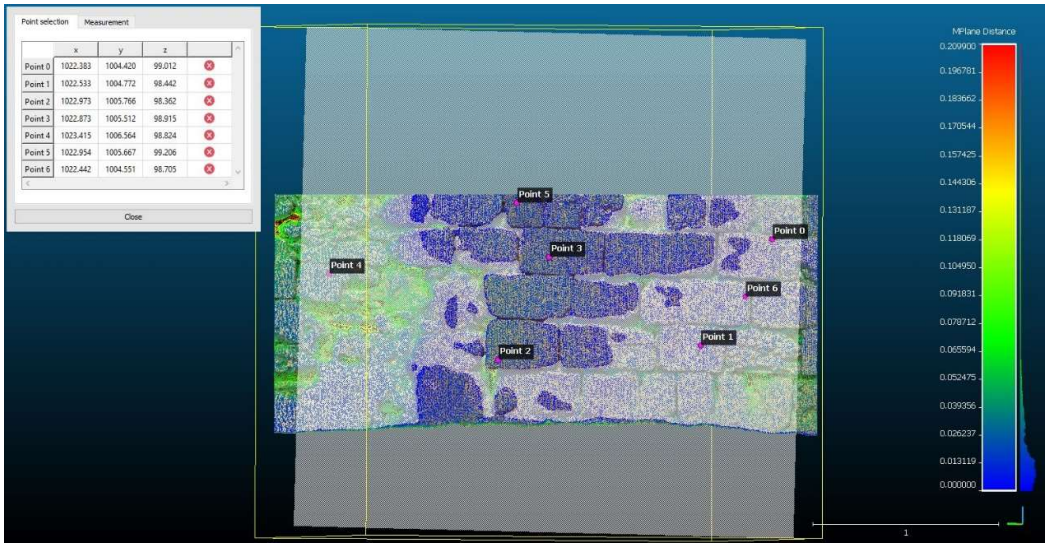

Figure 45: qMplane Plugin: Defining the best fit plane by picking points on the 3D point cloud.

The software automatically creates and applies a scalar field for the visualization of the normal distance between the point cloud and the plane leading to a first estimation of the highest and lowest degradation value. Further normal distance measurements against the plane were performed to acquire accurate and specific results about the depth of the degradation (Figure 46). This process leads to the estimation and evaluation of the material loss and the level of degradation for each part of the 4 Cultural Heritage buildings.

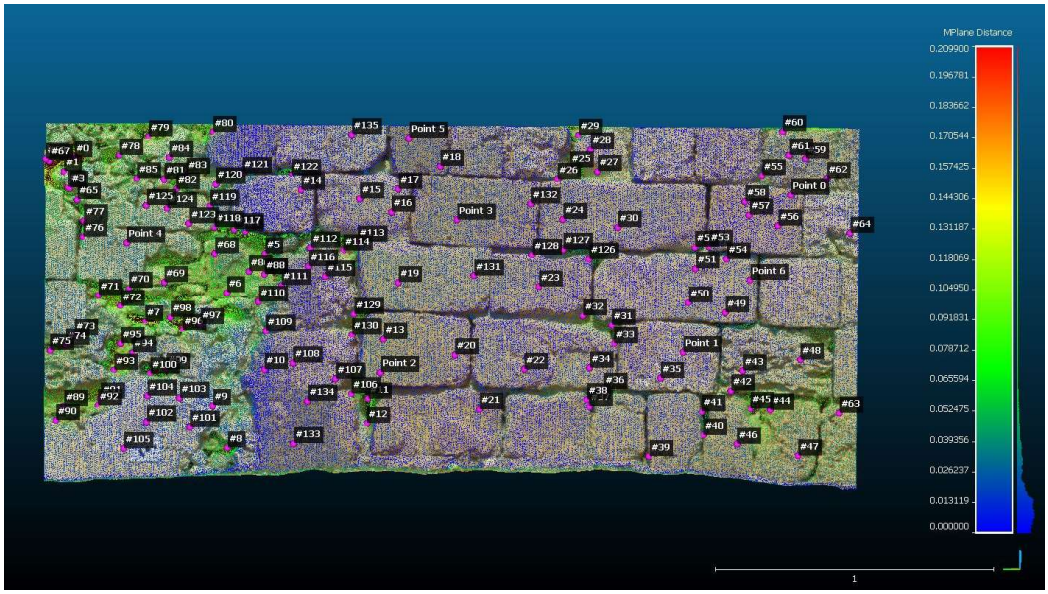

Figure 46: The selected points on the Detailed 3D Textured Model and the depth of the degradation.

The second approach for material loss estimation was based on the 'valley depth' algorithm (Conrad, 2012) of SAGA (SAGA, 2021), which is used for digital terrain models in order to estimate the depth of the valleys from a theoretical ridge elevation model. Valley depth is calculated as difference between the actual elevation and an interpolated ridge level. Ridge level interpolation uses the algorithm implemented in the 'Vertical Distance to Channel Network' tool of SAGA. It performs the following steps: (1) Definition of ridge cells using Strahler order (Stahler 1957) on the inverted DEM, (2) Interpolation of the ridge level, and (3) Subtraction of the original elevations from the ridge level.

Interestingly, this algorithm gives astonishingly good results for the material loss estimation when being used with a 3D model of a Cultural Heritage building. It only fails at the edges of the images or when other objects interfere with the wall plane. The result is derived very fast, without the need for any adjustments (apart maybe from the ridge order to be taken into consideration) and provides very accurate estimations (Figure 47).

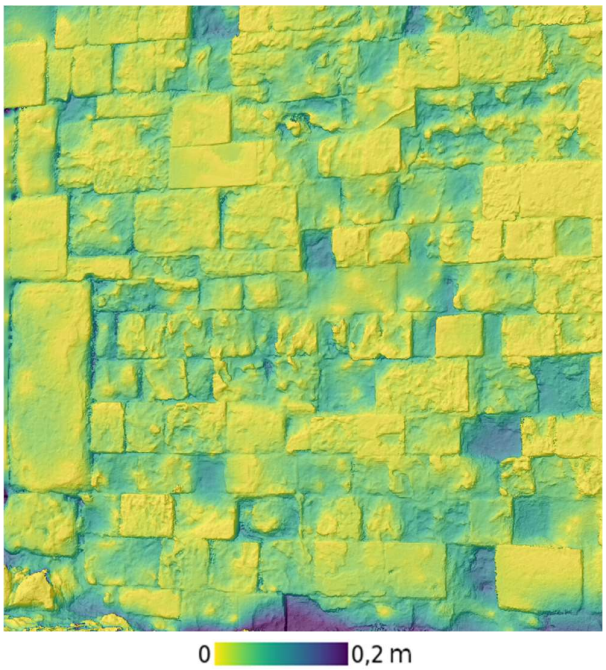

Figure 47: Material loss estimation using the valley depth algorithm (hillshade has been added as background for easier interpretation of the results)

Classification of material loss into desired classes can easily be done by applying appropriate thresholding conditions (Figure 48).

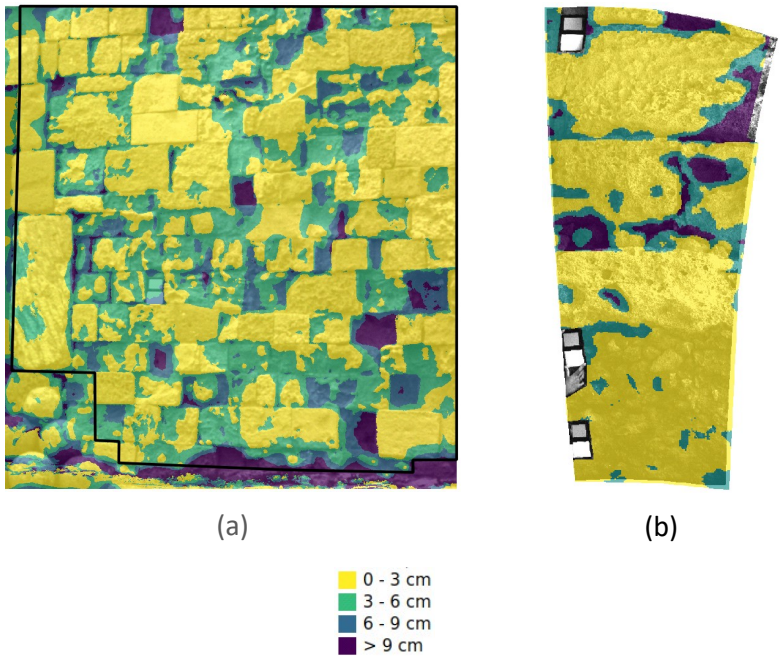

Figure 48: Valley depth classes for the 3D models of two wall parts of (a) the St Nikolas Fort and (b) the Roman Bridge.

4.4 Crack detection based on CNNs

Masonry structures represent the highest proportion of building stock worldwide, including Cultural Heritage assets and historical buildings. Currently, the structural condition of such structures and especially of monuments, is predominantly manually inspected, which is a laborious, costly and subjective process (Dais et al., 2021). With recent developments in computer vision and machine learning, there is a great opportunity to exploit RGB imagery to speed up this process in high precision and accuracy.

Computer vision for building pathologies detection has attracted the interest of researchers since quite some time. Vision-based crack detection is a non-destructive assessment technique, which can be useful especially for Cultural Heritage, where strict regulations apply and even simple interventions, such as placing crack-rulers, are not permitted by the conservation authorities (Dais et al., 2021). Recently, shallow and deep machine learning architectures applied on various types of imagery (RGB, multispectral etc.) are gaining ground, due to the increased automation and accuracy they offer.

In the context of HYPERION project, since an end-to-end method for semantically segment and map all the detected pathologies on the monuments is underway, a crack detection system is presented in this deliverable. The crack detection problem, is approached as an anomaly detection problem on RGB image patches, containing cracks or not. Towards this direction, a CNN (Convolutional Neural Network) system based on the VGG-16 network (Simonyan and Zisserman, 2014) is deployed and modified to classify the images or patches by delivering a specific class for the tested imagery; Crack or No crack and detect those cracks on the RGB imagery.

4.4.1 Related work

CNN based image classification can be categorized into three types: (I) image or image patch classification, (II) boundary box regression and (III) semantic segmentation (Zhang et al., 2019). In image classification, the image or image patch is labeled with a class. When boundary box regression is considered, a box bounds the detected object, that is a crack, and reveals its position and boundaries. To achieve this, the weights of the last dense layer are exploited. These two classification techniques have been extensively used to detect cracks and other defects delivering very promising results (Mohtasham et al., 2020, Feng et al., 2019, Cha et al., 2017, Cha et al., 2018. These techniques are implemented at block/patch level rather than at pixel level.

A combination of the above two classification types is performed for crack detection in the HYPERION project, since a class for the image or the image patch is predicted, delivering also the position of the cracks in the form of an overlayed heat map.

The exact location can be also provided in high accuracy using semantic segmentation methods. They can deliver the width or length of any defects/cracks since each pixel is assigned to a class label (Liu et al., 2019, Chen et al., 2018, Choi and Cha, 2020, LI et

al., 2018). Fully Convolutional Networks (FCNs), have been extensively used for semantic segmentation in many applications (Long et al., 2015). FCNs performed as an extended CNN where the final prediction was a semantically segmented image instead of a class identification (Dais et al., 2021). Recently, FCNs have been used widely for semantic segmentation on images containing cracks (Yang et al., 2018, Li et al., 2019, Hoskere et al, 2020, Ronneberger et al., 2015, Konig et al., 2019, Jekins et al., 2019, Liu et al., 2019, Ma et al., 2020).

Feature Pyramid Networks (FPNs) are typical model architectures to generate pyramidal feature representations for object detection. These architectures aim at extracting various features at different scales and then fuse them leading to pixel-level class predictions of higher accuracy (Lin et al., 2017). FPNs have also been used widely for crack detection (Zhang et al., 2018, Ni et al., 2020)).

Transfer learning has also been extensively implemented on different fields of computer vision with remarkable results and is considered suitable when the training dataset is small allowing for better performance and less computational effort. The intuition behind transfer learning for image classification is that if a model is trained on a large and general enough dataset, this model will effectively serve as a generic model of the visual world (Abadi et al., 2016). CNNs utilizing transfer learning have been used extensively for image classification and semantic segmentation of cracks (Yang et al, 2020, Bang et al., 2019, Li et al., 2019, Choi and Cha, 2020, Dais et al., 2021). Lately, different studies obtained remarkable results in crack segmentation by implementing region proposal networks followed by algorithms for pixel-level crack detection (Kalfarisi et al., 2020, Kang et al, 2020)

4.4.2 Dataset

To train and validate the CNN model, images with cracks are used from the test sites of the project. Specifically, square images of 224x224 pixels from Naillac, St. Nikolaos as well as from the roman bridge in Rodini (Rhodes) are used. Those images were split into two separate categories, facilitating the training and evaluation of the model later: “Cracks” and “No cracks”. Sample images are shown in the figures below where in Figure 49, images containing Cracks are presented while in Figure 50, images with No Cracks are shown. In total, only 50 images were used. Half of the images represent areas with various types and dimensions of cracks on various types of materials while the rest, represent areas without cracks, but complex enough. This small number of images with cracks, led us to implement the transfer learning approach presented in the next section.

For training the model, many approaches were tested and evaluated, including training on the 70% of the data and testing on the rest 30% of all the images shuffled, as well as training only on one test site and testing on another one, unseen by the model in the training process. This will later prove the generalization capacity of the trained models. To augment the available data during training, vertical and horizontal flipping, color jittering as well as random rotation was applied to the dataset.

Figure 49: Images of the sites containing Cracks.

Figure 50: Images of the sites with No Cracks.

4.4.3 The model

Main goal of the implemented CNN model here is the classification of the images into “Crack” and “No crack” as well as the important task of detection and localization of the cracks, if any, on the imagery. To achieve this dual purpose, the most efficient method consists in building a strong classifier. However, due to the small number of available imagery to train a full-scale model from scratch, this was decided to be achieved through a transfer learning approach.

Transfer learning consists of taking features learned on one problem, and leveraging them on a new, similar problem. In the presented case, features from a VGG16 CNN model (Simonyan and Zisserman, 2014) shown in Figure 51 trained on ImageNet (Deng et al., 2009) are used to kick-start a model meant to identify cracks.

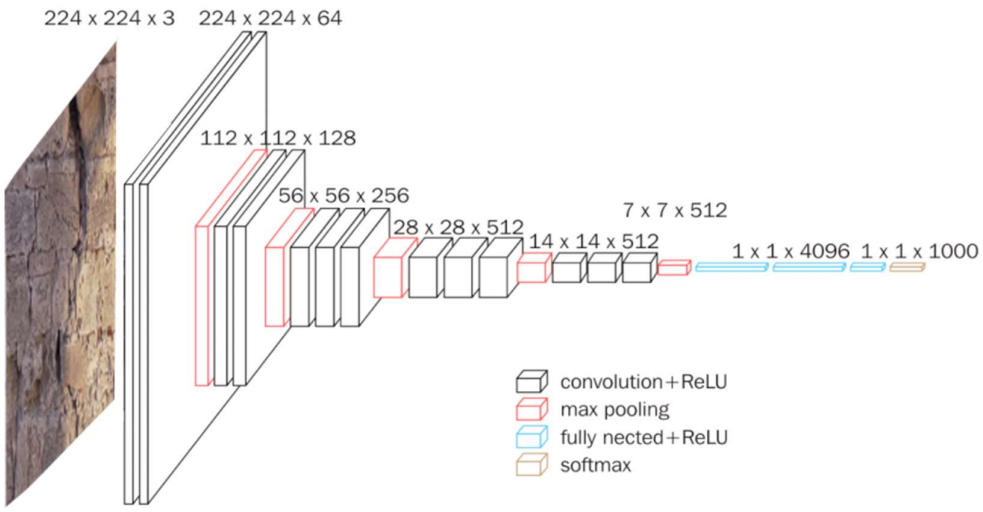

Figure 51: The VGG16 network architecture

The most common incarnation of transfer learning in the context of deep learning is the following workflow, which is also adopted here: Initially, layers from a previously trained model are taken and being frozen, to avoid destroying any of the information they contain during future training rounds. Consequently, some new, trainable layers on top of the frozen layers are added or/and some of the previous layers are being unfrozen. In the case presented here, the last four layers were unfrozen in order to turn the old features into predictions on a new dataset, such as the crack dataset here. To avoid overfitting, the top layers are being excluded from the original model and are replaced with a Global Average Pooling 2D layer which reduces the size of the preceding layer by taking the average of each feature map. In some cases, a last, optional step of fine-tuning is applied, which consists of unfreezing the entire model, and re-training it on the new data with a very low learning rate. A full description of the VGG16 network can be found in Simonyan and Zisserman (2014).

4.4.4 Results

To evaluate the different training and testing approaches, several metrics are used; *precision* which gives the ability of a classification model to return only relevant instances, *recall* which gives the ability of a classification model to identify all relevant instances, *F1 score* which is a single metric that combines *recall* and *precision* using the harmonic mean and *accuracy* which is the ratio of the correctly labeled subjects to the whole pool of subjects. While *recall* expresses the ability to find all relevant

instances in a dataset, *precision* expresses the proportion of the data points that the model says was relevant were actually relevant.

$$precision = \frac{TP}{TP + FP}, recall = \frac{TP}{TP + FN}$$

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

where TP are the true positives: data points labeled as “Crack” that are actually ”Cracks”, FP are the false positives: data points labeled as ”Cracks” that are actually ”No cracks”, TN are the true negatives: data points labeled as ”No cracks” that are actually ”No cracks” and FN are the false negatives: data points labeled as ”No cracks” that are actually ”Cracks”. In the tables below, the various results are presented. Those metrics are shown in Tables 3, 4, and 5.

Table 3:Accuracy metrics for the tests performed using the 30% of the shuffled data from all the three test sites, after training the model for 120 epochs on the rest 70% of the data.

	Precision	Recall	f1-score
Crack	1.00	0.86	0.92
No crack	0.88	1.00	0.93
Accuracy			0.93

Table 4: Accuracy metrics for the tests performed using the model trained on St. Nikolaos test site (60 epochs) over the Naillac test site data.

	Precision	Recall	f1-score
Crack	0.80	1.00	0.90
No crack	1.00	0.75	0.87
Accuracy			**0.89**

Table 5: Accuracy metrics for the tests performed using the model trained on Naillac test site (40 epochs) over the St. Nikolaos test site data.

	Precision	Recall	f1-score
Crack	0.80	1.00	0.90
No crack	1.00	0.75	0.87

Accuracy	**0.89**

Considering the small amount of available training, validation and testing data, the CNN delivered highly accurate results in terms of image/patch classification. It is shown that even when the model is trained using data from one test site and tested over another test site, resulted accuracy is very close to 90%, similar to the accuracies achieved in the vast majority of the state-of-the-rt methods.

4.4.5 Object Localization

To perform cracks localization on the imagery, many approaches can be used. The most common one is to replace the class score by bounding box location candidates. However, in the approach presented here, since a bounding box would contain greater areas of the image instead of only the detected crack or cracks, a heat map representation is used. This way, only the detected cracks are highlighted on the imagery in "red-ish" colour, providing also additional useful information. To achieve this, the weights of the final dense layer of the CNN are exploited. This activation map, is then bilinearly upsampled to have the same size as the original RGB image, and then it is projected on it, generating the images found in Figure 522 and Figure 53.

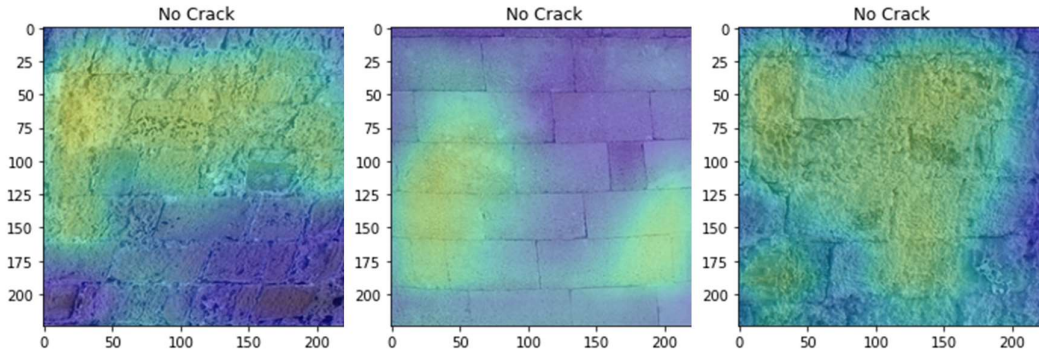

Figure 52: The resulted activation maps (detections) projected onto the tested RGB imagery, labeled also as "No Crack".

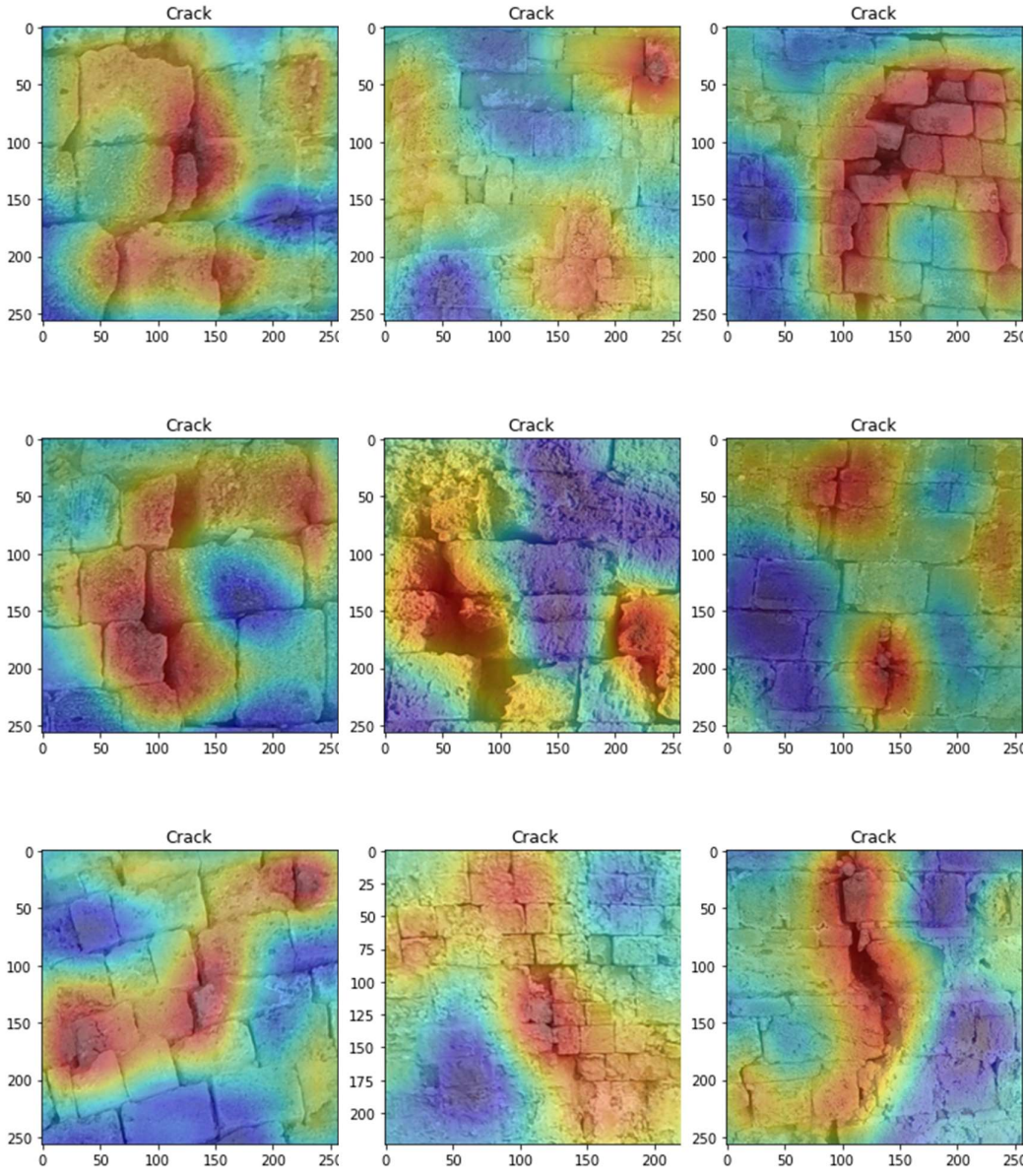

Figure 53: The resulted activation maps (detections) projected onto the tested RGB imagery, labeled also as "Crack".

Results suggested the high performance of the proposed approach, considering also the small numbers of epochs required for training. When training and testing was performed between different test sites, accuracy was slightly smaller compared to the one achieved when using the shuffled data, however this was expected due to the large variations between the cracks of the two sites. Nevertheless, those results met the accuracy delivered by more complex and computationally heavy approaches, requiring a large amount of data for training.

5. Conclusions

In this document, the specific methods and tools that have been developed within the HYPERION project for the routine monitoring are being reported and the routine monitoring products are presented.

Routine monitoring includes displacement and land cover change detection maps of the broad area for all the pilot sites (city of Rhodes, Granada, Venice and Tønsberg), three-dimensional models for all the CH assets that are studied within the Hyperion project, and deterioration and material loss estimation for specific parts of the facades of the Fort of Saint Nikolas and the Roman bridge in the Rhodes island.

Advanced methodologies using PS and SBAS functionalities and Convolutional Neural Network architectures have been developed and applied on satellite data, producing reliable land deformation and land cover change detection maps, respectively. Since monuments are not isolated in the geographical space, accurate assessment of changes in their broader area is important for preventing CH asset damages.

The described methods, tools and products of the 3D modelling process of the CH assets are crucial for both the routine monitoring and the post disaster monitoring of the HYPERION Project. Various techniques were explored for the 3D representation of the CH assets of the four pilot areas according to the difficulties, restrictions, complexity and size of each asset. Moreover, data from multiple sources were acquired and integrated in order to fully document the assets that will be used as reference data for the 4D representations. All these choices, methods and tools led to the required workflow and specifications that will be followed each time for the routine monitoring of the CH assets and after a disastrous event.

Finally, an integrated methodology based on 3D models and hyperspectral data has been successfully developed for providing deterioration and material loss maps. Data acquired by mobile hyperspectral sensors were used in order to build a Spectral Library for the CH asset materials, identify their deviations from normal situations and produce deterioration maps. SAGA-GIS functionalities have been used to provide material loss maps. Additionally, deep learning methods have been used to detect flaws and defects such as cracks on the monuments.

6. References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., ... & Zheng, X. (2016). Tensorflow: Large-scale machine learning on heterogeneous distributed systems. arXiv preprint arXiv:1603.04467.

Bang, S., Park, S., Kim, H., & Kim, H. (2019). Encoder–decoder network for pixel-level road crack detection in black-box images. Computer-Aided Civil and Infrastructure Engineering, 34(8), 713-727.

Boardman, J. W. 1998. “Leveraging the High Dimensionality of AVIRIS Data for Improved Subpixel Target Unmixing and Rejection of False Positives: Mixture Tuned Matched Filtering.” In Proceedings of the 5th JPL Geoscience Workshop, edited by R. O. Green, 55–56. Pasadena, CA: NASA Jet Propulsion Laboratory.

Boardman, J. W., F. A. Kruse, and R. O. Green. 1995.“Mapping Target Signatures via Partial Unmixing of AVIRIS Data” In Summaries of the Fifth Annual JPL Airborne Earth Science Workshop ,23–26, Washington, DC: JPL Publication 1

Calonder, Michael, Vincent Lepetit, Christoph Strecha, and Pascal Fua. "Brief: Binary robust independent elementary features." In European conference on computer vision, pp. 778-792. Springer, Berlin, Heidelberg, 2010.

Cerra, D., Plank, S., Lysandrou, V., & Tian, J. (2016). Cultural heritage sites in danger—towards automatic damage detection from space. *Remote Sensing*, *8*(9), 781.

Cha, Y. J., Choi, W., & Büyüköztürk, O. (2017). Deep learning-based crack damage detection using convolutional neural networks. Computer-Aided Civil and Infrastructure Engineering, 32(5), 361-378.

Cha, Y. J., Choi, W., Suh, G., Mahmoudkhani, S., & Büyüköztürk, O. (2018). Autonomous structural visual inspection using region-based deep learning for detecting multiple damage types. Computer-Aided Civil and Infrastructure Engineering, 33(9), 731-747.

Chen, L. C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2017). Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE transactions on pattern analysis and machine intelligence, 40(4), 834-848.

Chen, Jie, Ziyang Yuan, Jian Peng, Li Chen, Huang Haozhe, Jiawei Zhu, Yu Liu, and Haifeng Li. "DASNet: Dual attentive fully convolutional siamese networks for change detection of high resolution satellite images." IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing (2020).

Chen, Hao, and Zhenwei Shi. "A spatial-temporal attention-based method and a new dataset for remote sensing image change detection." Remote Sensing 12, no. 10 (2020): 1662.

Choi, W., & Cha, Y. J. (2019). SDDNet: Real-time crack segmentation. IEEE Transactions on Industrial Electronics, 67(9), 8016-8025.

Conrad O., 2012, SAGA-GIS Module Library Documentation (v2.3.0), Module Valley Depth, http://www.saga-gis.org/saga_tool_doc/2.3.0/ta_channels_7.html

Dais, D., Bal, I. E., Smyrou, E., & Sarhosis, V. (2021). Automatic crack classification and segmentation on masonry surfaces using convolutional neural networks and transfer learning. Automation in Construction, 125, 103606.

Dempster, Arthur P., Nan M. Laird, and Donald B. Rubin. "Maximum likelihood from incomplete data via the EM algorithm." Journal of the Royal Statistical Society: Series B (Methodological) 39, no. 1 (1977): 1-22.

Deng, Jia, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. "Imagenet: A large-scale hierarchical image database." In 2009 IEEE conference on computer vision and pattern recognition, pp. 248-255. IEEE, 2009.

Deng, J., Dong, W., & Socher, R. (2009). ImageNet: a large-scale hierarchical image database. IEEE Conf. on Computer Vision and Pattern Recognition.

Feng, C., Zhang, H., Wang, S., Li, Y., Wang, H., & Yan, F. (2019). Structural damage detection using deep convolutional neural network and transfer learning. KSCE Journal of Civil Engineering, 23(10), 4493-4502.

Fischler, Martin A., and Robert C. Bolles. "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography." Communications of the ACM 24, no. 6 (1981): 381-395.

François, Chollet. "Keras: The Python deep learning library." keras. io (2015).

Green, A. A., M. Berman, P. Switzer, and M. D. Craig. 1988. “A Transformation for Ordering Multispectral Data in Terms of Image Quality with Implications for Noise Removal.”, IEEE Transactions on Geoscience and Remote Sensing, 26: 65–74

Guo, Xiaoxin, Zhiwen Xu, Yinan Lu, and Yunjie Pang. "An application of Fourier-Mellin transform in image registration." In The Fifth International Conference on Computer and Information Technology (CIT'05), pp. 619-623. IEEE, 2005.

Hooper, A., H. Zebker, P. Segall, and B. Kampes, 2004. A new method for measuring deformation on volcanoes and other natural terrains using InSAR persistent scatterers, Geophysical Research Letters, 31(23), L23611

Hoskere, V., Narazaki, Y., Hoang, T. A., & Spencer Jr, B. F. (2020). MaDnet: multi-task semantic segmentation of multiple types of structural materials and damage in images of civil infrastructure. Journal of Civil Structural Health Monitoring, 10, 757-773.

Jenkins, M. D., Carr, T. A., Iglesias, M. I., Buggy, T., & Morison, G. (2018, September). A deep convolutional neural network for semantic pixel-wise segmentation of road and pavement surface cracks. In 2018 26th European Signal Processing Conference (EUSIPCO) (pp. 2120-2124). IEEE.

Jia, Yangqing, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, and Trevor Darrell. "Caffe: Convolutional architecture for fast feature embedding." In Proceedings of the 22nd ACM international conference on Multimedia, pp. 675-678. 2014.

ICOMOS-ISCS, 2011. "Illustrated glossary on stone deterioration patterns", International Council on Monuments and Sites International Scientific Committee for Stone. http://iscs.icomos.org/glossary.html

Kalfarisi, R., Wu, Z. Y., & Soh, K. (2020). Crack detection and segmentation using deep learning with 3D reality mesh model for quantitative assessment and integrated visualization. Journal of Computing in Civil Engineering, 34(3), 04020010.

Kang, D., Benipal, S. S., Gopal, D. L., & Cha, Y. J. (2020). Hybrid pixel-level concrete crack segmentation and quantification across complex backgrounds using deep learning. Automation in Construction, 118, 103291.

Kolokoussis P., Skamantzari M., Tapinaki S., Karathanassi V., Georgopoulos A. (2021). 3D and Hyperspectral data integration for assessing material degradation in medieval heritage buildings. Submitted and accepted for publication in the proceedings of the XXIV ISPRS Congress.

König, J., Jenkins, M. D., Barrie, P., Mannion, M., & Morison, G. (2019, September). A convolutional neural network for pavement surface crack segmentation using residual connections and attention gating. In 2019 IEEE International Conference on Image Processing (ICIP) (pp. 1460-1464). IEEE.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 25, 1097-1105.

Lebedev, M.A., Vizilter, Y.V., Vygolov, O.V., Knyaz, V.A. and Rubis, A.Y., 2018. CHANGE DETECTION IN REMOTE SENSING IMAGES USING CONDITIONAL ADVERSARIAL NETWORKS. International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences, 42(2).

Li, G., Ma, B., He, S., Ren, X., & Liu, Q. (2020). Automatic tunnel crack detection based on u-net and a convolutional neural network with alternately updated clique. Sensors, 20(3), 717.

Li, S., Zhao, X., & Zhou, G. (2019). Automatic pixel-level multiple damage detection of concrete structure using fully convolutional network. Computer-Aided Civil and Infrastructure Engineering, 34(7), 616-634.

Lin, T. Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2017). Feature pyramid networks for object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2117-2125).

Liu, Y., Yao, J., Lu, X., Xie, R., & Li, L. (2019). DeepCrack: A deep hierarchical feature learning architecture for crack segmentation. Neurocomputing, 338, 139-153.

Liu, Z., Cao, Y., Wang, Y., & Wang, W. (2019). Computer vision-based concrete crack detection using U-net fully convolutional networks. Automation in Construction, 104, 129-139.

Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3431-3440).

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, *60*(2), 91-110.

Mehra S. G., Ahadnejadb* V., Abbaspourc R. A., and Hamzehd M., 2013. "Using the mixture-tuned matched filtering method for lithological mapping with Landsat TM5 images.", International Journal of Remote Sensing, Vol. 34, No. 24, 8803–8816, http://dx.doi.org/10.1080/01431161.2013.853144, Taylor & Francis

Mohtasham Khani, M., Vahidnia, S., Ghasemzadeh, L., Ozturk, Y. E., Yuvalaklioglu, M., Akin, S., & Ure, N. K. (2020). Deep-learning-based crack detection with applications for the structural health monitoring of gas turbines. Structural Health Monitoring, 19(5), 1440-1452.

Ni, F., Zhang, J., & Chen, Z. (2019). Pixel-level crack delineation in images with convolutional feature fusion. Structural Control and Health Monitoring, 26(1), e2286.

Papagianni I., 2004. "Consolidation problems of Fortifications of the Medieval City of Rhodes", Research Project Report, Aristotelian University of Thessaloniki

Ronneberger, O., Fischer, P., & Brox, T. (2015, October). U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention (pp. 234-241). Springer, Cham.

Rosen, P.; Gurrola, E.; Sacco, G.F.; Zebker, H. The InSAR Computing Environment. In Proceedings of the EUSAR 2012 9th European Conference on Synthetic Aperture Radar, Nuremberg, Germany, 23–26 April 2012; pp. 730–733.

Rosten, Edward, and Tom Drummond. "Fusing points and lines for high performance tracking." In *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, vol. 2, pp. 1508-1515. IEEE, 2005.

Rublee, Ethan, Vincent Rabaud, Kurt Konolige, and Gary Bradski. "ORB: An efficient alternative to SIFT or SURF." In *2011 International conference on computer vision*, pp. 2564-2571. IEEE, 2011.

SAGA, 2021. System for Automated Geoscientific Analyses. University of Hamburg, Germany. http://www.saga-gis.org

Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv:1409.1556 (2014).

Strahler, A. N., 1957. "Quantitative analysis of watershed geomorphology", Transactions of the American Geophysical Union, 38 (6): 913–920, doi:10.1029/tr038i006p00913.

Tapinaki, S., Skamantzari, M., Chliverou, R., Evgenikou, V., Konidi, A. M., Ioannatou, E., Mylonas, A., & Georgopoulos A. (2019). 3D image based geometric documentation of a medieval fortress. International Archives of Photogrammetry, Remote Sensing and Spatial Information Science, XLII-2/W9, 699-705. https://doi.org/10.5194/isprs-archives-XLII-2-W9-699-2019

Tapinaki S., Skamantzari M., Anastasiou A., Koutros S., Syrokou E., Georgopoulos A. (2021). 3D Holistic documentation of heritage monuments in Rhodes. Submitted and accepted for publication in the archives of the 28th Symposium of CIPA.

Tompkins S., 1997. “Optimization of endmembers for spectral mixture analysis”, Remote Sens. Environ., vol. 59, no. 3, pp. 472–489.

Williams, A. P., and E. R. Hunt. 2002.“Estimation of Leafy Spurge Cover from Hyperspectral Imagery Using Mixture Tuned Matched Filtering.”, Remote Sensing of Environment, 82: 446–456

Xia, Gui-Song, Jingwen Hu, Fan Hu, Baoguang Shi, Xiang Bai, Yanfei Zhong, Liangpei Zhang, and Xiaoqiang Lu. "AID: A benchmark data set for performance evaluation of aerial scene classification." IEEE Transactions on Geoscience and Remote Sensing 55, no. 7 (2017): 3965-3981.

Yang, X., Li, H., Yu, Y., Luo, X., Huang, T., & Yang, X. (2018). Automatic pixel-level crack detection and measurement using fully convolutional network. Computer-Aided Civil and Infrastructure Engineering, 33(12), 1090-1109.

Yang, Q., Shi, W., Chen, J., & Lin, W. (2020). Deep convolution neural network-based transfer learning method for civil infrastructure crack detection. Automation in Construction, 116, 103199.

Yang, Zhuoqian, Tingting Dan, and Yang Yang. "Multi-temporal remote sensing image registration using deep convolutional features." IEEE Access 6 (2018): 38544-38555.

Yunjun, Z., Fattahi, H., & Amelung, F. (2019). Small baseline InSAR time series analysis: Unwrapping error correction and noise reduction. Computers & Geosciences, 104331.

Zhang, C., Chang, C. C., & Jamshidi, M. (2018). Bridge Damage Detection using a Single-Stage Detector and Field Inspection Images. arXiv preprint arXiv:1812.10590.

Zhang, J., Lu, C., Wang, J., Wang, L., & Yue, X. G. (2019). Concrete cracks detection based on FCN with dilated convolution. Applied Sciences, 9(13), 2686.

Zhang, Min, and Wenzhong Shi. "A feature difference convolutional neural network-based change detection method." IEEE Transactions on Geoscience and Remote Sensing 58, no. 10 (2020): 7232-7246.