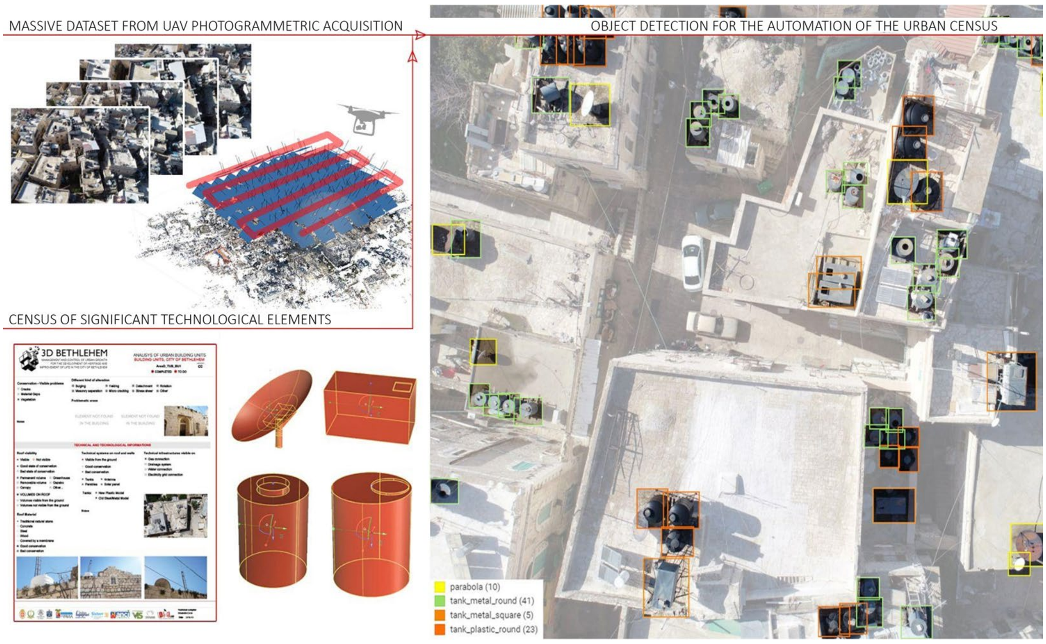
Object Detection Techniques Applied to UAV Photogrammetric Survey

Elisabetta Doria
Luca Carcano
Sandro Parrinello

Abstract

This project proposes an automated approach to the census of technological and architectural elements from massive photography datasets. This use case is built on photogrammetric close-range acquisitions performed via UAV over the roofs of the centre of Bethlehem, in order to map the water tanks for civilian use that create loads on historical buildings in a seismic area. The urban census was conducted within "3D Bethlehem. Management and control of urban growth for the development of Heritage and Improvement of life in the city of Bethlehem", a project promoted by AICS. The presented work leverages the project dataset to train Deep Learning models on a Cloud Infrastructure handling model lifecycle from training to deployment. Tests were conducted on historical buildings that show, among objects of interest, multiple spurious elements such as debris and junk. Such density creates complex scenarios for models that are trained to automate recurrent operations to assist large scale monitoring and management of the areas for different teams and municipalities.

Keywords
object detection, architectural census, urban monitoring, UAV photogrammetry, Bethlehem.



doi.org/10.3280/oa-845-c233

Introduction

The proposed research aims at automating the census process of technological and architectural elements in urban contexts, starting from massive photographic datasets. The use case this is built upon is that of installations on the covering of the historical centre of Bethlehem, selecting the dataset from UAV close-range photogrammetric acquisitions and fits within the scope of the International Cooperation Project *3D Bethlehem. Management and control of urban growth for the development of Heritage and Improvement of life in the city of Bethlehem* [1]. Starting from 2018 researchers from DAda-LAB, from University of Pavia, conducted multiple surveys to build a complete documentation of the historical centre of Bethlehem. This research leverages close-range photogrammetric acquisitions [Picchio 2019] consisting of more than 9000 UAV photographs of the city centre.
While the final goal of the *3D Bethlehem* was that of providing the municipality with a digital tool to improve governance of the urban complex of Bethlehem, the presented piece of work arises from the need of an automated tool for the monitoring of specific elements of the urban environment. The analytical process to understand urban contexts is tied to the construction of reliable data bases to develop strategies for communication and virtual fruition of spaces. Historical centres retain information concerning the numerous events that impacted the city over the years and create an heterogeneous and variable dataset that can be systematised [Bocconcino 2019]. The chosen use case stems from the need to map the water tanks that civilians place on the roofs. Tanks are used as water supplies but create structural criticalities placing localised loads on the coverages in a seismic environment [2]. The urban landscape follows the morphology of underlying hills and is dotted with a high number of tanks that are visible from multiple angles: mapping these elements is one of the intervention priorities for

Fig. 1. Different tanks visible on the covering. The majority of surveyed tanks has a capacity of either 500 or 1000 litres, causing high load on the coverings.

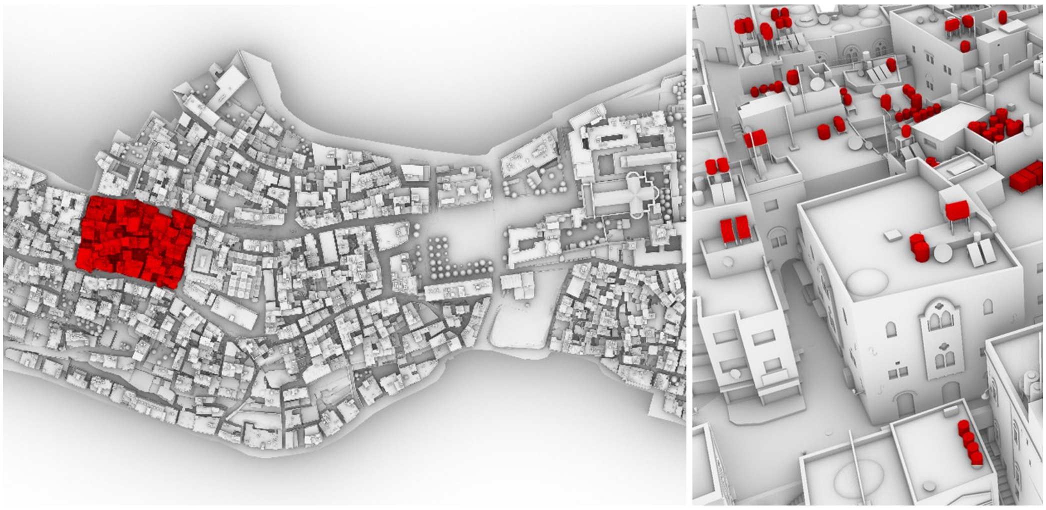


Fig. 2. Three-dimensional mesh model of the historic center of *3D Bethlehem* project. Left: Experimental Area R highlighted with respect to the context of the historic center of Bethlehem. Right: cisterns highlighted on the buildings in Area R to understand the high number of elements present on the roofs.

the municipality. This requirement fits in the scope of UNESCO site management plan, in which to develop alternative proposals for the replacement of such infrastructures. Tanks fall in different categories: plastic tanks are the most widespread (54%) and are on average well preserved but have a heavy impact due to colour, either black or white, contrast with the building; metal tanks (46%) on the other side are either cylindrical or box shape and in a worse state of conservation (Fig. 1).
Across the four years span of the *3D Bethlehem* a census of such elements was performed and tank data was structured in a database [3]. During provisioning, the tanks that show flow control issues show leaks on coverings and facades causing material pathologies, vegetation growth and creating moist areas on the walls. Tanks are visible all around Bethlehem, both in recently built structures (that show neither additional elements stored nor debris) and in historical buildings that show complex morphology along with stacked debris and scrap.

Dataset Acquisition Methodology

Studying the historical city centre of Bethlehem was conducted from data surveyed between 2018 and 2021 to detect different building fabrics detectable in the urban context. During the surveying campaign a series of interventions aimed at consolidating and preventing structural damage have been detected across different neighbourhoods. *In situ* studies of Bethlehem city centre led to a subdivision in 20 areas based on historical subdivisions and road patterns across the city. Areas have been progressively named with letters from A (Church of Nativity neighbourhood) to V and different states of conservation can be found across different Areas; newly constructed buildings (V) can be found as well as stratified historical neighbourhoods (Area F) and only few of these Areas have been targeted by projects for renovation and consolidations [4]. Al-Anatreh (Area F), the historical christian neighbourhood was widely impacted by interventions between 2003 and 2005 [Nasser 2005], while the other areas saw targeted interventions on specific building units having high historical and monumental value [5]. Despite interventions for consolidation and rebuilding of the urban fabric there is a widespread portion of built fabric that show the presence of ruins, especially in P and R areas (Al Hreizat e Al Fawagreh Quarter) area U (Al Farahieh Quarter), C and N (Al Najajreh Quarter). This research focuses on Area R (highlighted in Fig. 3), a neighbourhood characterised by high building density with heterogeneous coverings considering both the typology and state of conservation, with some of those showing the presence of debris that hinder automated recognition of objects since partial occlusion may occur [Saleh 2021]. To build the experimental setup the team leveraged the aerial photography dataset to train Deep Learning models on a cloud infrastructure. Photographs from Area R (389 pictures) have been imported and from there a subset was manually tagged highlighting the objects to be detected by the Object Detection algorithm.

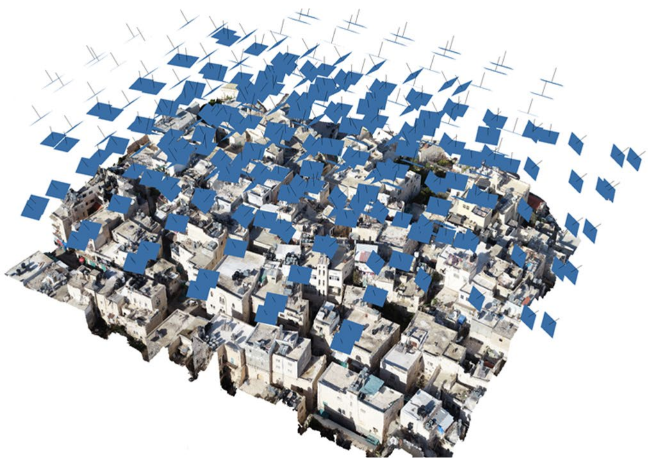
Photographic Acquisition via UAV

A fundamental phase of the proposed research was to have a vast photographic archive of close-range coverings available. For the roofs of the historic center of Bethlehem, over 9,000 photographs were acquired to cover an area of over 260,000 square meters, split into neighbourhoods. The photographs were taken with different camera angles:
– zenith photographs of the roofs;
– photographs angled from four different sides that make up the neighbourhood (Fig. 3).
The choice of taking the photos from different angles was necessary in order to generate an SfM model of the roofs, from which it was possible to orient and extrapolate ortho-images [Picchio 2020]. The photographs have an overlap of 65-80% between them. The photographs allowed to obtain a very varied dataset both as urban extension and as a variety of points of view from which the cisterns were photographed [6].
As can be seen in Fig. 3, there are critical issues related not only to environmental conditions such as lights and shadows but also to the deposits of objects present on the roofs that make the urban fabric particularly complex. The goal of the project is that of training an automated recognition system able to detect tanks on the roofs, subsequently feeding a programmed monitoring process. Such monitoring allows to verify the compliance with a series of criteria set by the municipality of Bethlehem such as:
- The reduction of the visual perception of the cisterns from the roads;
- Their positioning in safe points and without structural criticalities.

As regards the structural verification of the criticality due to the presence of cisterns, was carried out a cross-reading of the data with what was acquired in the building census [7].

Deep Learning for Object Detection

Object detection is a branch of computer vision aimed at identifying objects in images. The task is twofold since the algorithm has to isolate the object of interest in a potentially complex scene and then correctly identify the isolated object with the correct class.
Since the introduction of Convolutional Neural Networks [Lecun 1995], Deep Learning (DL) models have become the standard thanks to the high performances achieved [Redmon 2016]. DL models propose a layer based approach in which the image is fed into a series of processing stages that extract features that are then fed to a classifier.
The computational complexity of this problem grows with the number of images in the training dataset and size in pixels of the images. In the simplest case, a sliding window is applied to each portion of the image and slid across the entire frame (convolution) to be subsequently processed by stages deeper in the model architecture. These supervised models require large datasets to provide a wide number of examples of the objects to be identified and multiple executions to train the models via backpropagation.



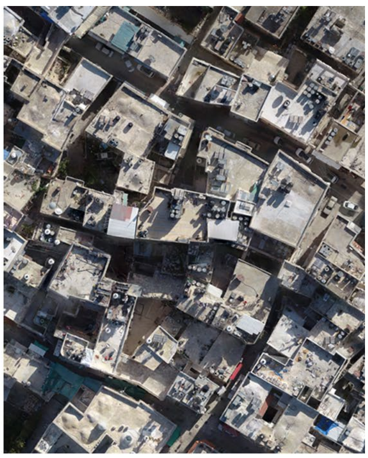


Fig. 3. Example of SfM photogrammetric acquisition of an area with photographs taken at 50 m from the height of the roof from which the survey was conducted. Right: extraction of the ortho image from the photogrammetric model.

The dataset is manually tagged identifying objects of interest, the data is then processed, matched with the ground truth provided and weights that govern the model behaviour are corrected to match the computed and expected result. Execution is repeated until a plateau in performance or a target number of training epochs is reached. With the availability of computing power and acceleration techniques, the complexity of models could grow and new methodologies for learning stemmed from this growth in capabilities. AutoML is a family of models, ensemble of models and feature engineering approaches that leverage ensemble and transfer learning techniques that allow faster construction of ML models and push towards the democratisation of Artificial Intelligence [He 2021].

Cloud Computing – Integrated Machine Learning Platforms

This use case consists of a massive initial dataset and a potentially ever growing one, therefore computational complexity sets a constraint on the available approaches. Abstracting from the computational power available on premises removed the boundaries set by the physical hardware shifting the focus on the problem itself [Rivera 2020]. On premise set ups need to be expanded by buying new hardware if the dataset grows or a new choice of algorithm stresses the existing hardware configuration. Cloud Platforms offer on demand resources in a pay per use fashion, allowing to tailor the needs to the as-is state of the problem and scaling up or down on demand, depending on the computing power and storage required. The presented infrastructure was implemented within the Google Cloud Platform (GCP) set of tools. GCP is a suite of cloud computing tools that offer infrastructure, platform and software as a service tools running on Google data centres. Data storage is performed on the cloud itself allowing for high reliability, fast transfer speeds and resilience to failure. This use case should be considered as a piece of a wider project for surveying, documentation and feeding of an informative system for cultural heritage; this infrastructure provides storage, computing resources, machine learning management, dashboarding and other tools in a unified platform for the entire project lifecycle. This use case addresses a single area of the centre of Bethlehem, but the platform can be scaled to accommodate larger datasets, more complex models or additional components. A Cloud based infrastructure grows and shrinks as a tailored environment built around each of the use cases of the project. The components of this project are: Google Cloud Storage (GCS) buckets to store images, predictions, trained models, and Vertex AI. Vertex AI is a machine learning platform that covers the entire ML lifecycle from dataset tagging and splitting to training, deployment, performance tracking and management of ML models in a serverless environment. The models are accessible via UI and programmatically, and in case a new model outperforms the previous one it can be effortlessly swapped without concerns for the end users. An AutoML model was trained on a subset of 70 UAV images from the aerial photography dataset.

Fig. 4. Figure shows a ortho image with the output of the model overlayed on top. Highlighted examples of Correct Detections, mismatches, competing labels and missed objects. Both images show high precision.Left: False Negative: missed plastic tank close to metallic tanks.
Right: False Negative: missed plastic tank. Black tanks area critical due to the shadows on the roofs.

Results

Initial tests were conducted on a 30 image dataset (UAV photographs), but the model lacked in the recall department failing to identify the correct group of pixels to be classified.
The higher the precision, the fewer false positives predicted, the higher the recall, the fewer false negatives, or the fewer predictions missed. With a threshold set at 0.5:

- 80.5% precision;
- 28.0% recall.

Increasing the size of the training dataset vastly improved performance, to achieve current results the team used 70 UAV photographs for training. Increasing the size of the training dataset vastly improved performance since each image contains tens or hundreds of objects. Tagging a higher number of images increased by an order of magnitude the number of training examples. The second training instance was run on:

- N° parabola: 497
- N° tanks metal_round: 3180
- N° tanks metal_square:695
- N° tanks plastic_round: 188514

Portions of orthoimages of the same area were used for validation. UAV pictures captured from different angles and distances provide a natural data augmentation for the algorithm (Fig. 5). The retrained model, threshold at 0.5:

- 90.8% precision;
- 61.0% recall.

The greatest issues with recall are in dark areas, in which low light reduces contrast between objects and the high amount of cluttering and occlusion on the roofs where wood planks and debris are placed on the covers and on the objects themselves. (Fig. 6)
Since this model is intended as a screening tool to identify criticalities and plan interventions and maintenance, when deploying it for use the team opted for higher tolerance (hence lower threshold) accepting a higher number of false positives but ensuring a higher recall.

Average precision	0.266
Precision	80.8%
Recall	28%
Created	1 Aug 2021, 20:48:58
Total images	30
Training images	24
Validation images	4
Test images	2

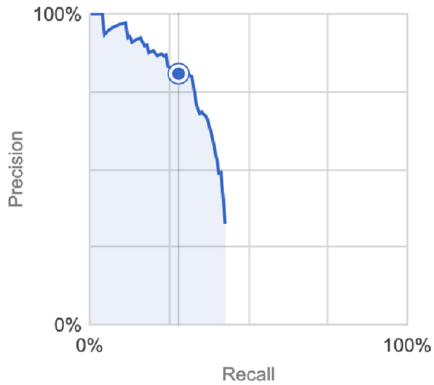


Fig. 5. Precision/recall curve - 30 image dataset.

Average precision	0.583
Precision	90.8%
Recall	61.8%
Created	12 Feb 2022, 21:18:30
Total images	85
Training images	71
Validation images	7
Test images	7

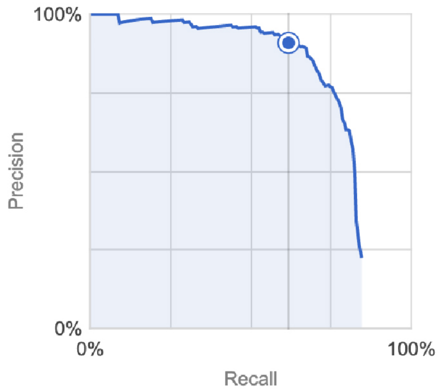


Fig. 6. Precision/recall curve - 70 image dataset.

Conclusions and Future Developments

The research will broaden its scope, extending object detection to different areas of the city to study how the urban setting can alter the results of tests. The algorithm will be evaluated both on morphologically complex historic neighbourhoods and on recent constructions that show a disaggregated morphology, more isolated buildings and regular shapes. This second phase will allow the team to evaluate whether the algorithms should be retrained to handle the entire area of the city. Different elements can be detected and mapped alongside the tanks as well other characteristics of the buildings. Satellite dishes were included in the first phase of the experiments, providing interesting results in both precision and recall. Repeating the same analyses at different moments of time, devising an inspection schedule, will allow to compare the results and map the evolution of the built elements through time, identifying criticalities. State updates are performed thanks to ordinary periodical activities, planned within a management schedule, and extraordinary activities in case of critical events that can vary timings and needs for the inspections.

The field of heritage conservation is tightly bound to best practises for fruition and safeguarding of buildings. Architectural conservation can stem from diagnosing the state of conservation of heritage, within the context it is located, anticipating the need for a procedural and systematic vision [Cecchi 2006]. Management and conservation of cultural heritage, also considering sets of actions that get scheduled and coordinated through time (Intervention time schedule), aims at improving the quality and identity of the elements. Such projects require multidisciplinary contributions, economical and management evaluations, as well as the involvement of the local population in the process of recognising the value and development opportunities of the existing heritage [Della Torre 2008]. The presented research targets the automation of processes in a widespread and complex built environment. Managing such scenario is a challenge for municipalities and automation facilitates the monitoring activities for entities and administrations. This process, if extended to the recognition of different architectural and technological elements, allows to improve the planning of protocols for the preventive maintenance tied to the actions context, maximising enhancement goals for the built heritage.

Notes

[1] *3D Bethlehem. Management and control of urban growth for the development of heritage and improvement of life in the city of Bethlehem* is a cooperation project promoted by AICS, the Italian Agency for Cooperation and Development. The project is coordinated by the Municipality of Pavia, with a partnership made up of the Municipality of Bethlehem, the University of Pavia (scientific coordination), the University of Bethlehem, the Province of Pavia, the Order of Engineers of the Province of Pavia, the SISTERR territorial system of Pavia for international cooperation. APS, ANCI Lombardia, VIS – International Voluntary Service for NGO Development and Palestinian Engineers Association – Jerusalem Centre. The project is scientifically coordinated by Prof. Sandro Parrinello and DAda-Lab laboratory of the DICAr – Department of Civil Engineering and Architecture of the University of Pavia.

[2] The Dead Sea fault is an area subject to the risk of major seismic events. See in particular the seismic risk map produced by An-Najah National University – Nablus.

[3] The GIS-connected building database consists of structural and technological information about the buildings. The percentages relating to the different types of tanks present were obtained from the overall reading of the cards in the GIS system. For an in-depth study on the research relating to the database on the historic center of Bethlehem see: Doria Elisabetta, Picchio Francesca, 2020.

[4] The data refers to interventions mapped during the urban census up to 2019.

[5] Examples of localized interventions are the Syriac Hosh in Hreizat Quarter (2011-2013) in area P; Al-Badd Museum in Najajreh Quarter (2014) in area M and the refurbishment of the road network in area U (2018-2019) and Star Street in area O (2019).

[6] The photogrammetric acquisition of the roofs was conducted by Ph.D. Francesca Picchio during the *3D Bethlehem* project. For specifics on the methods of acquisitions conducted and the tools used, see: Parrinello Sandro, Francesca Picchio (2019).

[7] This research was enforced in a collaboration between DJI Enterprise and the University of Pavia for the development of research activities, and the promotion of the different ways of using drones for cultural heritage. This collaboration is based on the "Agreement for the development of research activities about the digital documentation of cultural heritage and landscape using drones" between the Department of Civil Engineering and Architecture of the University of Pavia and iFlight Technology Company Limited, signed in February 2020, lasting three years.

References

Bocconcino Maurizio Marco, Manzone Fabio (2019). Sistemi informativi e strumenti grafici per la manutenzione di manufatti complessi – Information systems and graphic tools for the maintenance of complex buildings. In Garda Emilia, Piantanida Paolo, Mele Caterina (eds.). *Colloqui.AT.e 2019. Ingegno e costruzione nell'epoca della complessità (Torino, 25-27 settembre 2019)*. Torino: Politecnico di torino, pp. 679-688.

Cecchi Roberto (2006). *I Beni culturali. Testimonianza materiale di civiltà*. Milano: Spirali Edizioni.

Ćosović Marijana, Janković Radmila (2020). CNN Classification of the Cultural Heritage Images. In *2020 19th International Symposium INFOTEH-JAHORINA (INFOTEH)*. Piscataway: IEEE pp. 1-6.

Dabdoub Nasser, Ḥifẓ al-Turāth al-Thaqāfī (2005). *Anatreh Quarter: An Urban and Architectural Study of a Bethlehem Quarter*. Bethlehem: Centre for Cultural Heritage Preservation.

Della Torre Stefano (2008). Quadro generale di riferimento per l'analisi dei bisogni del comparto dei Beni culturali. In *Skill. Rivista semestrale di ENAIP Lombardia*, 35, 2008, pp. 31-28.

De Marco Raffaella, Miceli Alessia, Parrinello Sandro (2020). An assessment on morphological survey calibration and the automation of digital drawing for the reliable documentation and conservation analysis of out-of-scale buildings. In *2020 IMEKO TC-4 International Conference on Metrology for Archaeology and Cultural Heritage (Trento, 22-24 October 2020*). Budapest: IMEKO-International Measurement Federation Secretariat, pp. 226-231.

Doria Elisabetta, Picchio Francesca (2020). Il censimento della città: il rilievo tecnologico e la strutturazione di banche dati per il patrimonio edilizio. In Parrinello Sandro (ed.). *3D Bethlehem. Sistema di gestione e controllo della crescita urbana per lo sviluppo del patrimonio e miglioramento della vita nella città di Betlemme,* vol.1. Firenze: Edifir Edizioni, pp. 151-172.

He Xin, Zhao Kaiyong, Chu Xiaowen (2021). AutoML: A survey of the state-of-the-art. In *Knowledge-Based Systems*, 212, 2021, 106622.

Janković Radmila (2020). Machine Learning Models for Cultural Heritage Image Classification: Comparison Based on Attribute Selection. In *Information,* 11 (1), 12, pp. 1-13.

Lakshmanan Valliappa (2018). *Data Science on the Google Cloud Platform*. Paris: Editions O'Reilly.

Lecun Yann, Bengio Yoshua (1998). Convolutional networks for Images, Speech, and Time-Series. In *The Handbook of Brain Theory and Neural Networks*, october 1998, pp. 255-258.

Montella Massimo (2009). *Valore e valorizzazione del patrimonio culturale storico*. Milano: Mondadori.

Parrinello Sandro, Picchio Francesca (2019). Integration and comparison of close-range SfM methodologies for the analysis and the development of the historical city center of Bethlehem. In *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.,* XLII-2/W9, 2019, pp. 589-595.

Parrinello Sandro, Picchio Francesca, Barazzoni Paola, De Marco Raffaella, Doria Elisabetta (2019). 3D Bethlehem, un progetto di cooperazione internazionale per la conoscenza della città stratificata = 3D Bethlehem, an international cooperation project for the knowledge of the stratified city. In Belardi Paolo (ed.). *UID 2019 - Riflessioni: l'arte del disegno/il disegno dell'arte / Reflections: the art of drawing/the drawing of art*. Roma: Gangemi, pp. 899-908.

Picchio Francesca (2019). La fotogrammetria close range per la costituzione di modelli SfM affidabili sul sistema urbano. In Parrinello Sandro (ed.). *3D Bethlehem. Sistema di gestione e controllo della crescita urbana per lo sviluppo del patrimonio e miglioramento della vita nella città di Betlemme,* vol. 1. Firenze: Edifir Edizioni, pp. 101-117.

Picchio Francesca (2020). Acquisition protocols for UAV photogrammetric data. Comparison in methodological SfM procedures from architectural till urban scale. In Barba Salvatore, Dell'Amico Anna, Limongello Marco, Parrinello Sandro (ed.). *D-SITE. Drones – Systems of Information on culTural hEritage. For a spatial and social investigation*. Pavia: Pavia University Press, pp. 70-79.

Redmon Joseph, Divvala Santosh, Girshick Ross, Farhadi Ali (2016). You Only Look Once: Unified, Real-Time Object Detection. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Piscataway: IEEE, pp. 779-788.

Rivera Juan (2020). Object detection with a model trained in Google Cloud AutoML. In *Practical TensorFlow.js*. Berkeley: Apress, pp. 163-184.

Saleh Kaziwa, Szénási Sándor, Vámossy Zoltán (2021). Occlusion Handling in Generic Object Detection: A Review. In *2021 IEEE 19th World Symposium on Applied Machine Intelligence and Informatics (SAMI)(21-23 January 2021)*. Herl'any: IEEE, pp. 477-484.

Authors
Elisabetta Doria, Dept. of Civil Engineering and Architecture, University of Pavia, elisabetta.doria@unipv.it
Luca Carcano, Dept. of Civil Engineering and Architecture, University of Pavia, luca.carcano@gmail.com
Sandro Parrinello, Dept. of Civil Engineering and Architecture, University of Pavia, sandro.parrinello@unipv.it

 Isbn 9788835141945