

BOLLETTINO SIFET

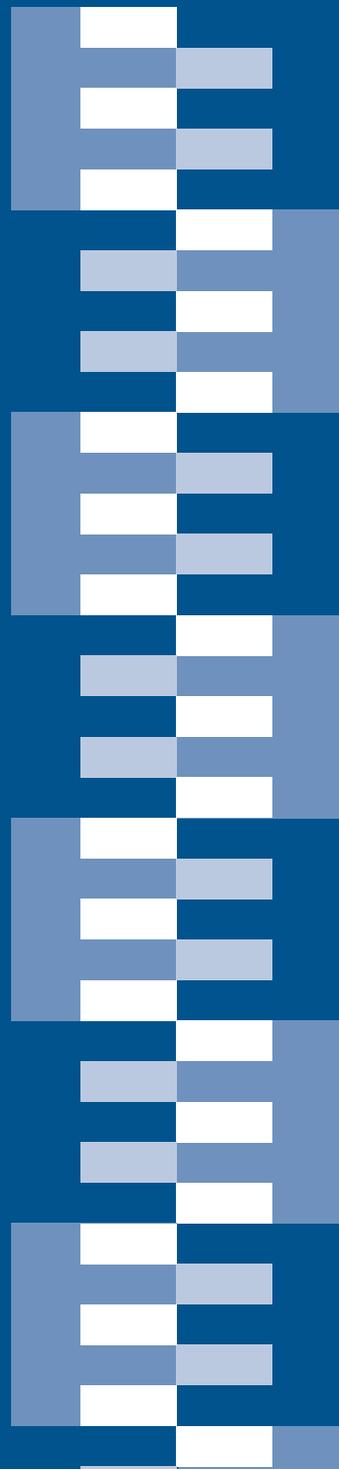
ANNO 2020

ISSN 2784-9473

BOLLETTINO

SIFET

SOCIETÀ ITALIANA DI FOTOGRAMMETRIA E TOPOGRAFIA



20

Bollettino della Società Italiana di Fotogrammetria e Topografia

ISSN 1721-971X & ISSN 2784-9473 (online)



Direttore responsabile | Chief Editor

Prof. Andrea Maria Lingua

Segreteria di Redazione | Publishing Editor

e-mail: redazione@sifet.org

I contributi di tutte le sezioni sono sottoposti a revisione tra pari, se accettati sono condivisi con licenza CC-BY-SA, il copyright rimane all'autore(i).

All the editorial sections are open submissions, peer-reviewed and indexed, copyright remains with the author(s).



Segmentazione semantica delle nuvole di punti utilizzando tecniche di apprendimento profondo per il patrimonio culturale

Roberto Pierdicca, Marina Paolanti, Francesca Matrone, Massimo Martini, Christian Morbidoni, Eva Malinverni, Emanuele Frontoni, Andrea Lingua

pg.. 1-9

Un benchmark per la segmentazione semantica di nuvole di punti di beni culturali

Francesca Matrone, Andrea Lingua, Roberto Pierdicca, Eva Malinverni, Marina Paolanti, Eleonora Grilli, Fabio Remondino, Arnadi Muriyoso, Tania Landes

pg.. 10-18

Intelligenza artificiale e realtà aumentata per la condivisione del patrimonio culturale

Roberta Spallone, Valerio Palma

pg.. 19-26

Apprendimento da dati sintetici di nuvole di punti per la segmentazione semantica di edifici storici

Christian Morbidoni, Roberto Pierdicca, Marina Paolanti, Raissa Mammoli, Ramona Quattrini

pg.. 27-37

Approccio gerarchico di machine learning per la segmentazione semantica di nuvole di punti 3D

Eleonora Grilli, Simone Teruggi, Francesco Fassi, Fabio Remondino, Michele Russo

pg.. 38-46

Ontologie geografiche nel dominio spaziale urbano e del patrimonio costruito

Antonia Spanò, Elisabetta Colucci

pg.. 47-56

SEGMENTAZIONE SEMANTICA DELLE NUVOLE DI PUNTI UTILIZZANDO TECNICHE DI APPRENDIMENTO PROFONDO PER IL PATRIMONIO CULTURALE

POINT CLOUD SEMANTIC SEGMENTATION USING A DEEP LEARNING FRAMEWORK FOR CULTURAL HERITAGE

R. Pierdicca^a, M. Paolanti^b, F. Matrone^c, M. Martini^b, C. Morbidoni^b, E.S. Malinverni^a, E. Frontoni^b, A.M. Lingua^c

^a Dipartimento di Ingegneria Civile, Edile e dell'Architettura, Università Politecnica delle Marche, 60100 Ancona, Italy; r.pierdicca@univpm.it (R.P.); e.s.malinverni@univpm.it (E.S.M.)

^b Dipartimento di Ingegneria dell'Informazione, Università Politecnica delle Marche, 60100 Ancona, Italy; m.martini@pm.univpm.it (M.M.); c.morbidoni@univpm.it (C.M.); e.frontoni@univpm.it (E.F.)

^c Dipartimento di Ingegneria dell'Ambiente, del Territorio e delle Infrastrutture, Politecnico di Torino, 10129 Torino, Italy; francesca.matrone@polito.it (F.M.); andrea.lingua@polito.it (A.M.L.)

PAROLE CHIAVE: segmentazione semantica, patrimonio culturale digitale, nuvole di punti, apprendimento profondo

KEY WORDS: Semantic segmentation, Digital Cultural Heritage, point clouds, Deep Learning

RIASSUNTO

Nell'ambito del patrimonio culturale digitale (*Digital Cultural Heritage*), la segmentazione semantica delle nuvole di punti 3D con tecniche di apprendimento profondo (*Deep Learning*) può aiutare a riconoscere automaticamente elementi architettonici storici a un adeguato livello di dettaglio. Questo permetterebbe di accelerare il processo di modellazione dei modelli HBIM (*Historical Building Information Modeling*), a partire dai dati di rilievo. In questo lavoro viene proposto un *framework* innovativo di DL per la segmentazione delle nuvole di punti che, a partire da una rete convoluzionale dello stato dell'arte (*Dynamic Graph Convolutional Neural Network*) ottimizza il processo di segmentazione semantica grazie all'introduzione di *features* significative come normali e componente radiometrica. Per testare tale approccio, sono state utilizzate alcune nuvole di punti facenti parte di un nuovo dataset pubblicamente disponibile: l'ArCH (*Architectural Cultural Heritage*) dataset. Questo dataset comprende 17 nuvole di punti annotate, derivanti dall'unione di più scansioni singole o dall'integrazione di queste ultime con rilievi fotogrammetrici. Le scene coinvolte sono sia interne che esterne, con chiese, cappelle, chiostri, portici e logge costituiti da elementi architettonici molto differenti tra loro. Queste nuvole di punti appartengono a differenti periodi storici con diversi lessici, in modo da rendere il dataset il meno possibile uniforme e omogeneo (nella ripetizione degli elementi architettonici) e i risultati il più generalizzabile possibile. Dagli esperimenti effettuati, la nuova rete (DGCNN-Mod) fornisce elevati livelli di accuratezza, dimostrando l'efficacia dell'approccio proposto.

ABSTRACT

In the Digital Cultural Heritage (DCH) domain, the semantic segmentation of 3D Point Clouds with Deep Learning (DL) techniques can help to recognize historical architectural elements, at an adequate level of detail, and thus speed up the process of modeling of historical buildings for developing BIM models from survey data, referred to as HBIM (Historical Building Information Modeling). In this paper, we propose a DL framework for Point Cloud segmentation, which employs an improved DGCNN (Dynamic Graph Convolutional Neural Network) by adding meaningful features such as normal and colour. The approach has been applied to a newly collected DCH Dataset which is publicly available: ArCH (Architectural Cultural Heritage) Dataset. This dataset comprises 17 labeled points clouds, derived from the union of several single scans or from the integration of the latter with photogrammetric surveys. The involved scenes are both indoor and outdoor, with churches, chapels, cloisters, porticoes and loggias covered by a variety of vaults and beared by many different types of columns. They belong to different historical periods and different styles, in order to make the dataset the least possible uniform and homogeneous (in the repetition of the architectural elements) and the results as general as possible. The experiments yield high accuracy, demonstrating the effectiveness and suitability of the proposed approach.

1. INTRODUZIONE

Nell'ambito del patrimonio culturale digitale (*Digital Cultural Heritage* - DCH), la generazione di nuvole di punti 3D è, al giorno d'oggi, uno tra i modi più efficienti per gestire le risorse e i dati del patrimonio culturale (Cultural Heritage - CH). La rappresentazione dei beni culturali attraverso dati 3D permette di svolgere diversi compiti: analisi morfologica, mappatura del degrado o arricchimento e informatizzazione dei dati sono, infatti, solo alcuni esempi dei possibili modi per sfruttare una rappresentazione virtuale così ricca di informazioni. La gestione delle informazioni nel DCH è fondamentale per una migliore

interpretazione dei dati sul patrimonio e per lo sviluppo di strategie di conservazione appropriate. Una strategia di gestione delle informazioni efficiente dovrebbe prendere in considerazione tre concetti principali: classificazione, organizzazione delle relazioni gerarchiche e arricchimento semantico (Grilli et al., 2018). Le scansioni laser e la fotogrammetria digitale (*Close Range Photogrammetry* - CRP) consentono di generare grandi quantità di scene 3D dettagliate, con informazioni geometriche dipendenti dal metodo impiegato. Inoltre, lo sviluppo negli ultimi anni di tecnologie come il *Mobile Mapping System* (MMS) sta contribuendo alla massiccia documentazione metrica 3D del patrimonio costruito (Masiero et

al., 2018; Bronzino et al., 2019). Pertanto, la gestione, l'elaborazione e l'interpretazione delle nuvole di punti sta acquisendo sempre più importanza nel campo della geomatica e della rappresentazione digitale. Queste strutture geometriche stanno diventando progressivamente obbligatorie non solo per la creazione di esperienze multimediali (Barazzetti et al., 2015; Osello et al., 2018), ma anche (e principalmente) per supportare il processo di modellazione 3D (Balletti et al., 2016; Bolognesi and Garagnani, 2018; Chiabrando et al., 2016; Fregonese et al., 2017), dove anche le reti neurali stanno iniziando ad essere sempre più impiegate (Barazzetti and Previtali, 2019; Borin and Cavazzini, 2019).

Allo stesso tempo, le recenti tendenze di ricerca nell'ambito dell'*Historical Building Information Modeling* (HBIM) sono finalizzate alla gestione di molteplici tipologie di dati del patrimonio architettonico (Bitelli et al., 2017; Bruno and Roncella, 2018; Oreni et al., 2017;), tra le quali si affronta anche il problema della trasformazione dei modelli 3D da una rappresentazione geometrica a un "contenitore" di dati arricchito e informativo (Quattrini et al., 2017). Il raggiungimento di tale risultato non è banale, poiché i modelli HBIM sono generalmente basati su processi *scan-to-BIM* che permettono di generare un modello 3D parametrico a partire dalla nuvola di punti (Capone et al., 2019). Questi processi, sebbene molto affidabili poiché realizzati manualmente da esperti del settore, presentano due ostacoli: in primo luogo, richiedono molto tempo e, in secondo luogo, si basano su un'elevata quantità di dati, derivanti dalle nuvole di punti (sia da TLS che da CRP), che contengono molte più informazioni di quelle richieste per descrivere un oggetto parametrico.

La letteratura dimostra che, fino ad ora, i metodi tradizionali applicati all'ambito DCH fanno ancora ampio uso di operazioni manuali per interpretare gli oggetti del patrimonio dalle nuvole di punti (Murtiyoso and Grussenmeyer, 2019; Grilli et al., 2019; Spina et al., 2011). A tal fine, ultimamente, un campo di ricerca molto promettente è lo sviluppo di *framework* basati su tecniche di *Deep Learning* (DL) per le nuvole di punti. Ne sono un esempio alcune reti neurali come PointNet o PointNet++ (Qi et al., 2017a; Qi et al., 2017b) che forniscono metodologie più potenti ed efficienti per gestire i dati 3D (Wang et al., 2018). Tali sistemi sono programmati per assolvere a tre compiti principali: classificazione e segmentazione semantica sia di oggetti che di parti di essi (*part segmentation*) a partire dalle nuvole di punti. La classificazione delle nuvole di punti prende l'intero dato come *input* e fornisce in *output* la classe di appartenenza dell'*input* iniziale. La segmentazione mira invece a classificare ogni punto in una parte specifica della nuvola di punti (Zhang et al., 2019). Sebbene la letteratura sulla segmentazione delle istanze 3D sia limitata, se confrontata con quella 2D (a causa dell'elevata memoria e costi computazionali richiesti dalla rete neurale convoluzionale (CNN) per la comprensione della scena (Song and Xiao, 2016; Ma et al., 2019)), questi *framework* possono facilitare il riconoscimento di elementi architettonici storici, a un adeguato livello di dettaglio, accelerando così il processo di ricostruzione delle geometrie nell'ambiente BIM (Tang et al., 2010; Tamke et al. 2016; Macher et al., 2017; Thompson and Boehm, 2015). Questi metodi, oggi, non sono ancora stati pienamente sfruttati e applicati per il riconoscimento automatico degli elementi appartenenti al patrimonio architettonico, sebbene le nuove potenze di calcolo e le peculiarità delle nuove reti 3D diano risultati molto promettenti, che meritano di essere indagati e approfonditi. Infatti, anche se si sono rivelate adatte alla gestione di nuvole di punti con forme regolari, i beni culturali sono caratterizzati da geometrie complesse, molto variabili tra loro e pienamente definibili solo con un alto livello di dettaglio. In (Grilli et al., 2019), gli autori studiano il potenziale offerto dagli approcci DL per la classificazione supervisionata del

patrimonio 3D ottenendo risultati promettenti. Tuttavia, il lavoro non fa fronte alla natura irregolare dei dati del mondo CH e mostra limiti nella generalizzazione dei metodi.

Per far fronte a questo problema, il presente articolo propone un innovativo *framework* basato sul DL per la segmentazione semantica delle nuvole di punti, ispirato al lavoro presentato in (Wang et al., 2019). Invece di impiegare punti singoli come PointNet (Qi et al., 2017a), l'approccio proposto in (Wang et al., 2019) sfrutta le strutture geometriche locali costruendo un grafo dei vicini locali e applicando operazioni di convoluzione sui bordi che collegano coppie di punti adiacenti.

Il contributo di questo lavoro di ricerca consiste nel miglioramento di tale rete (Dynamic Graph Convolutional Neural Network – DGCNN), alla quale vengono aggiunte caratteristiche rilevanti del dato come il colore RGB e codificato HSV. Gli esperimenti sono stati eseguiti su un set di dati DCH completamente nuovo, nel quale sono state selezionate 11 nuvole di punti annotate, derivate dall'unione di più scansioni o dall'integrazione di queste ultime con rilievi fotogrammetrici.

Un quadro complessivo del *framework* sviluppato è riportato nella Figura 1. Tale flusso di lavoro potrebbe rappresentare una linea guida per ulteriori esperimenti da parte di altri ricercatori che si occupano di segmentazione semantica delle nuvole di punti con approcci DL applicati al settore dei beni culturali.

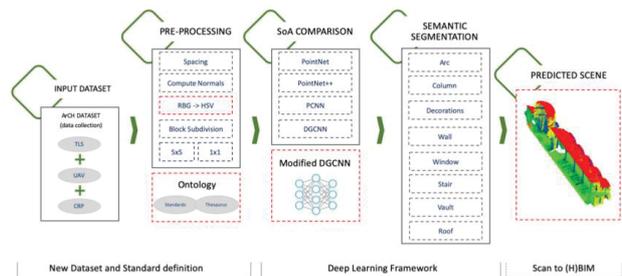


Figura 1. Workflow per la segmentazione semantica delle nuvole di punti.

I principali contributi di questo lavoro rispetto agli approcci dello stato dell'arte si possono quindi così riassumere: (i) creazione di un *framework* DL per la segmentazione semantica DCH delle nuvole di punti, utile per la documentazione 3D di monumenti e siti culturali; (ii) sperimentazione e implementazione di un approccio DL basato sulla DGCNN con *features* aggiuntive; (iii) primi test su un nuovo dataset DCH che è pubblicamente disponibile per la comunità scientifica, al fine di testare e confrontare diversi metodi e iv) definizione di un insieme coerente di classi di elementi architettonici, basato sull'analisi delle classificazioni degli standard esistenti.

Il documento è organizzato come segue. La sezione 2 fornisce una descrizione dei criteri adottati per la segmentazione semantica delle nuvole di punti del patrimonio culturale. La sezione 3 descrive l'approccio proposto, che modifica una rete DL per l'apprendimento dalle nuvole di punti, in modo da sfruttare al meglio le *features* geometriche locali delle nuvole di punti, e il nuovo dataset testato per il dominio DCH. La sezione 4 offre un'ampia valutazione comparativa e un'analisi dettagliata dell'intera metodologia. Infine, nelle sezioni 5 e 6 si traggono le conclusioni e sono discusse le direzioni future per questo campo di ricerca.

2. STATO DELL'ARTE

In questa sezione è analizzata la letteratura relativa alla classificazione e segmentazione semantica per la rappresentazione digitale del patrimonio culturale.

Ad oggi, sono molti gli studi in cui le nuvole di punti vengono utilizzate per il riconoscimento e la ricostruzione di geometrie relative a modelli BIM (Tamke et al., 2016; Macher et al., 2017; Thomson and Boehm, 2015), tuttavia questi metodi non sono stati ancora applicati al DCH e non sfruttano pienamente le strategie proprie del DL. I beni culturali sono infatti caratterizzati da geometrie più complesse, molto variabili anche all'interno della stessa classe e descrivibili solo con un alto livello di dettaglio, rendendo quindi molto più complicato applicare metodi DL a questo ambito. Nonostante esistano già alcuni lavori che classificano le immagini DCH impiegando diversi tipi di tecniche (Mathias et al., 2011; Oses et al., 2014; Stathopoulou and Remondino, 2019; Llamas et al., 2017), sono ancora poche le ricerche che sfruttano direttamente le nuvole di punti di beni culturali architettonici per la classificazione o la segmentazione semantica tramite tecniche di intelligenza artificiale (Grilli and Remondino, 2019). Uno di questi è (Barsanti et al., 2017), dove viene proposta una segmentazione di modelli 3D di edifici storici per l'analisi FEA, a partire da nuvole di punti e mesh. Gli autori hanno testato alcuni algoritmi come quello di *region growing*, direttamente sulle nuvole di punti, dimostrandone l'efficacia per la segmentazione di strutture piane e ben definite; tuttavia, geometrie più complesse come curve o parti lacunose non sono state segmentate correttamente e i tempi di calcolo sono aumentati notevolmente. Sono stati poi testati alcuni software per la segmentazione diretta delle mesh, ma in questo caso i risultati mostrano che il processo è ancora completamente manuale e, nell'unico esempio in cui la segmentazione è semiautomatica, il software non è in grado di gestire grandi modelli, quindi è necessario suddividere il file e procedere all'analisi delle singole porzioni. Infine, è da sottolineare che il caso studio utilizzato per la segmentazione delle mesh (il tempio di Nettuno a Paestum) è un'architettura piuttosto regolare e simmetrica, quindi relativamente facile da segmentare sulla base di alcuni piani orizzontali.

Poiché le nuvole di punti sono strutture geometriche di natura irregolare, caratterizzate dalla mancanza di una griglia, con un'elevata variabilità di densità e non ordinate (Zaheer et al., 2017), sfruttare l'intelligenza artificiale per automatizzare il processo di riconoscimento automatico delle geometrie è stimolante sotto molti aspetti.

Dalle ricerche analizzate, l'unico tentativo recente di utilizzare il DL per la classificazione semantica delle nuvole di punti di DCH è il lavoro di Grilli et al. (2019). Il metodo descritto consiste in un flusso di lavoro composto dall'estrazione e selezione di *features* in grado di descrivere i diversi elementi architettonici, facilitando così la classificazione automatica (Weinmann et al., 2015) grazie all'utilizzo di strategie di Machine Learning (ML). Tra tutti i classificatori, gli autori utilizzano il Random Forest (RF) e il One-vs.-One, ottimizzando i parametri per il RF e scegliendo quelli con il valore di *F1-score* più alto in *Scikit-learn*. In questo modo si ottengono ottime prestazioni nella maggior parte delle classi individuate, tuttavia non è stata effettuata alcuna correlazione delle *features* e, soprattutto, quest'ultime vengono selezionate in base alle peculiarità del caso di studio da analizzare. D'altra parte, per gli approcci DL, usano una CNN 1D e una 2D, oltre alla Bi-LSTM RNN (*Recurrent Neural Network*) che viene solitamente utilizzata per la predizione di sequenze o testi. La scelta di questo tipo di reti neurali è dovuta all'interpretazione della nuvola di punti come sequenza di punti e, in questo modo, i risultati del ML superano quelli del DL. Ciò potrebbe essere dovuto alla scelta di non utilizzare nessuna delle recenti reti progettate appositamente per tenere conto della terza dimensione dei dati della nuvola di punti. Inoltre, la fase di test del DL viene svolta sulla restante parte della nuvola di punti, molto simile ai dati presentati in fase di addestramento della rete,

pertanto questa impostazione non permette di generalizzare adeguatamente la metodologia proposta.

3. MATERIALI E METODI

In questa sezione è introdotto il *framework* DL illustrato in Figura 1 e il dataset utilizzato per la valutazione. Si è scelto di utilizzare la DGCNN modificata per la segmentazione semantica delle nuvole di punti di beni culturali. Ulteriori dettagli sono forniti nelle seguenti sottosezioni. La metodologia viene valutata su parte dell'ArCH dataset, oggi pubblicamente disponibile online e appositamente creato per questo lavoro.

3.1 ArCH dataset per la segmentazione semantica della nuvola di punti

Nello stato dell'arte, i dataset più utilizzati per addestrare le reti neurali sono: ModelNet 40 (Wu et al., 2015) con più di 100 K modelli CAD di oggetti, principalmente arredamenti, di 40 categorie diverse; KITTI (Geiger et al., 2013) che include immagini e scansioni laser per la navigazione autonoma; Sydney Urban Objects acquisito con Velodyne HDL-64E LiDAR in ambienti urbani con 26 classi e 631 scansioni individuali; Semantic3D (Hackel et al., 2017) con scene urbane come chiese, strade, ferrovie, piazze e così via; S3DIS (Armeni et al., 2016) che comprende principalmente uffici ed è stato acquisito con lo scanner Matterport con sensori a luce strutturata 3D e, infine, Oakland 3-D Point Cloud (Munoz et al., 2009) costituito da nuvole di punti 3D annotate, acquisite da un laser scanner mobile in un ambiente urbano. La maggior parte dei dataset attuali raccoglie dati da ambienti urbani, con scansioni composte da circa 100 K punti e, ad oggi, non ci sono ancora dataset pubblicati incentrati su beni culturali immobili con un adeguato livello di dettaglio.

Le nuvole di punti impiegate per i seguenti test fanno parte del più ampio ArCH (*Architectural Cultural Heritage*) dataset (Matrone et al., 2020a). Dal momento che il lavoro qui presentato è stato sviluppato prima della presentazione ufficiale di questo dataset (composto da un totale di 17 nuvole di punti annotate) sono state utilizzate solo quelle scene che erano al momento disponibili, ossia 11 nuvole di punti (Figura 2).

Le scene coinvolte sono di ambienti sia interni che esterni, con chiese, cappelle, chiostri, portici e loggiati. Appartengono a differenti periodi storici e differenti lessici architettonici, in modo da rendere il dataset il meno uniforme o omogeneo possibile (nella ripetizione degli elementi architettonici) e i risultati il più generali possibile. A differenza di molti dataset esistenti, è stato annotato manualmente da esperti di dominio, fornendo così un dataset più accurato.

I casi studio qui presi in esame sono: alcune cappelle dei Sacri Monti di Ghiffa (SMG) e Varallo (SMV); il Santuario del Trompone (TR) in provincia di Vercelli; la Chiesa di Santo Stefano (CA) in provincia di Torino e la scena interna del Castello del Valentino (VA) a Torino:

- I Sacri Monti di Ghiffa e Varallo. Questi due complessi devozionali dell'Italia settentrionale sono stati inseriti nel 2003 nella Lista del Patrimonio Mondiale dell'UNESCO (WHL). Nel caso del Sacro Monte di Ghiffa è stata scelta una loggia di 30 m con colonne e mezze lesene in pietra toscana; mentre per il Sacro Monte di Varallo sono stati inseriti nel dataset 6 edifici, contenenti un totale di 16 cappelle, alcune delle quali molto complesse dal punto di vista architettonico: volte a botte, talvolta lunettate, volte a crociera, portici, balaustre.
- Il Santuario del Trompone (TR). Si tratta di un ampio complesso risalente al XVI secolo ed è costituito da una chiesa (40x10 m circa) e da un chiostro (25x25 m circa),

entrambi inclusi nel dataset. La struttura interna della chiesa è composta da 3 navate coperte da volte a crociera sorrette a loro volta da colonne in pietra. Vi è inoltre un'ampia cupola absidale e una serie di lesene che ricoprono le pareti laterali.

- La Chiesa di Santo Stefano (CA) ha una struttura compositiva completamente diversa rispetto alla precedente, essendo una piccola chiesa campestre dell'XI secolo. Sono presenti una muratura in pietra, non intonacata, archi in mattoni sopra le finestrelle e una fascia lombarda che definisce una modanatura decorata sotto il tetto di tegole.
- La scena interna del Castello del Valentino (VA) è una sala aulica facente parte di un edificio storico del XVII secolo. Questa sala è coperta da una volta a crociera poggianti su sei robuste colonne in breccia. Ampie portefinestre illuminano la stanza e nicchie ovali contornate da stucchi decorativi sono collocate sulle pareti laterali. Anche questo caso studio fa parte di un sito seriale inserito nella lista dell'UNESCO dal 1997.

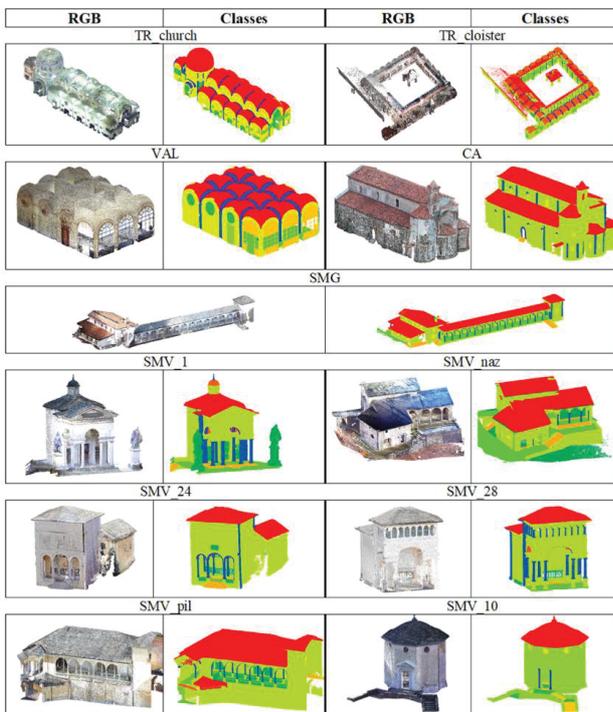


Figura 2. Scene dell'ArCH dataset utilizzate per i test. Nella colonna di sinistra le nuvole di punti RGB e in quella a destra le scene annotate. Sono state identificate 10 classi: Arco, Colonna, Porta, Pavimento, Tetto, Scale, Volta, Muro, Finestra e Decorazione. La classe Decorazione comprende tutti i punti non assegnati alle classi precedenti, come panchine, balaustre, dipinti, altari e così via.

Nella maggior parte dei casi, la scena finale è stata ottenuta mediante l'integrazione di diverse nuvole di punti, quelle acquisite con il laser scanner terrestre (TLS), e quelle derivanti dalla fotogrammetria (prevalentemente aerea per il rilievo dei tetti). Questa integrazione si traduce in una nuvola di punti completa, con densità diversa a seconda dei sensori utilizzati, che tuttavia porta ad aumentare la dimensione complessiva della nuvola di punti e richiede una fase di pre-elaborazione per la rete neurale. La struttura comune delle nuvole di punti è quindi basata sulla sequenza delle coordinate x , y , z e dei valori R , G , B .

3.2 Pre-elaborazione dei dati

Per predisporre il dataset per la rete sono state effettuate operazioni di pre-elaborazione al fine di rendere più omogenee le strutture della nuvola. I metodi di pre-elaborazione, per questo dataset, hanno seguito 3 passaggi: traslazione spaziale, sottocampionamento e scelta delle *features*.

La traslazione spaziale delle nuvole di punti è necessaria a causa della georeferenziazione delle scene: i valori delle coordinate sono infatti troppo grandi per essere elaborati dalla rete neurale, quindi le coordinate vengono troncate e ogni singola scena viene spostata spazialmente vicino all'origine del sistema (0,0,0). Questa operazione da un lato ha portato alla perdita della georeferenziazione, dall'altro ha però permesso di ridurre la dimensione dei file e lo spazio da analizzare, portando così anche a una diminuzione della potenza di calcolo richiesta.

Più complessa è stata invece l'operazione di sottocampionamento, resa necessaria a causa dell'elevato numero di punti (per lo più ridondanti) presenti in ogni scena (> 20M punti). È stato infatti necessario stabilire quale delle tre diverse opzioni di sottocampionamento fosse la più adeguata per fornire la migliore tipologia di dati di *input* alla rete neurale. L'opzione del sottocampionamento casuale è stata scartata perché limita la ripetibilità del test, quindi sono stati testati entrambi gli altri due metodi: *octree* e *space*. Il primo è efficiente per l'estrazione del punto più vicino (*nearest neighbour*), mentre il secondo fornisce, nella nuvola di punti di *output*, punti non più vicini di una distanza specificata. Per quanto riguarda lo spazio è stato impostato uno spazio minimo tra i punti di 0,01 m, in questo modo è garantito un elevato livello di dettaglio, ma allo stesso tempo è possibile ridurre notevolmente il numero di punti e la dimensione il file, oltre a regolarizzare la struttura geometrica della nuvola di punti. Per quanto riguarda *octree*, applicato solo nelle prime prove su metà della scena della Chiesa del Trompone, è stato impostato il livello 20, in modo che il numero dei punti finali fosse più o meno simile a quello della scena sottocampionata con il metodo spaziale. Il software utilizzato per questa operazione è *CloudCompare*.

Per quanto riguarda invece l'estrazione di *features* direttamente dalle nuvole di punti, ci si addentra in un campo di ricerca aperto e in continua evoluzione. La maggior parte delle *features* sono estratte a mano per compiti specifici e possono essere suddivise e classificate in intrinseche ed estrinseche, o utilizzate anche per descrittori locali e globali (Weinmann et al., 2015). Le caratteristiche locali definiscono le proprietà statistiche delle informazioni geometriche del vicinato locale, mentre le caratteristiche globali descrivono l'intera geometria della nuvola di punti. Quelle maggiormente utilizzate nel mondo del DL sono quelle locali, come i descrittori basati sugli autovalori (*eigenvalue*) o i *3D Shape context*, tuttavia in questo caso caso, poiché le ultime reti sviluppate (Qi et al., 2017b; Wang et al., 2019) tendono a lasciare che la rete stessa apprenda le *features* e poiché l'obiettivo principale è generalizzare il più possibile, oltre a ridurre il coinvolgimento umano nelle fasi di pre-elaborazione, le uniche *features* calcolate sono le normali. Quest'ultime sono state calcolate su *CloudCompare* e orientate con impostazioni diverse a seconda del modello di superficie e del sistema di acquisizione dati 3D. Nello specifico è stato utilizzato un "modello di superficie locale" piano o quadrico come approssimazione della superficie per il calcolo delle normali ed è stato impostato un *minimum spanning tree* con $KNN = 10$ per il loro orientamento. Quest'ultimo è stato ulteriormente verificato su MATLAB®.

3.3 Deep learning per la segmentazione semantica delle nuvole di punti

Le reti neurali profonde nello stato dell'arte sono specificamente progettate per affrontare l'irregolarità delle nuvole di punti, gestendo direttamente i dati grezzi piuttosto che utilizzare una rappresentazione regolare intermedia (metodi basati sulla creazione di un set di immagini dalle nuvole di punti, definiti come *multi-view*, o sulla loro rasterizzazione, definiti come *voxel-based*).

In questo contributo vengono quindi confrontate e poi valutate le prestazioni ottenute con le scene dall'ArCH dataset di alcune architetture dello stato dell'arte rispetto alla rete DGCNN modificata qui proposta. Le reti neurali selezionate sono:

- PointNet (Qi et al., 2017a), pioniere di questo approccio, che garantisce l'invarianza della permutazione dei punti operando su ogni punto in modo indipendente e applicando una funzione simmetrica per accumulare le *features*;
- la sua evoluzione PointNet ++ (Qi et al., 2017b) che analizza punti vicini preferendo agire su ciascuno separatamente, e consente lo sfruttamento delle caratteristiche locali anche se con ancora alcune importanti limitazioni;
- PCNN (Atzmon et al., 2018), un *framework* DL per l'applicazione della CNN alle nuvole di punti in cui sono coinvolti gli operatori di estensione e restrizione che consentono l'utilizzo delle funzioni volumetriche associate alla nuvola di punti;
- DGCNN (Wang et al., 2019) che risolve queste carenze aggiungendo l'operazione *EdgeConv*, un modulo che descrive le relazioni tra un punto e i suoi vicini (Figura 3). Questo modulo è invariante alle permutazioni ed è in grado di raggruppare i punti grazie al grafo locale, apprendendo dai bordi che collegano i punti.

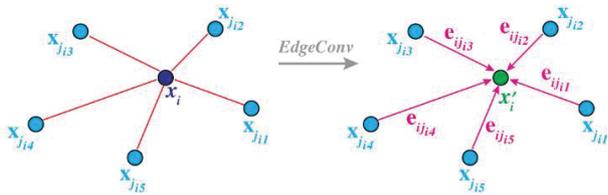


Figura 3. Modulo *EdgeConv*. L'output di *EdgeConv* viene calcolato aggregando le *features* del bordo associate a tutti i bordi provenienti da ciascun vertice connesso.

3.4 DGCNN per il dataset delle nuvole di punti DCH

Negli esperimenti qui descritti, ci si basa sull'implementazione della DGCNN fornita da (Wang et al., 2019). Tale implementazione utilizza *k-Nearest Neighbor* (kNN) per individuare i *k* punti più vicini al punto da classificare, definendo così la regione adiacente del punto. Basandosi sulla teoria dei grafi, $G = (V, E)$ dove V sono i vertici ed E sono i bordi (*edges*), le caratteristiche del bordo vengono calcolate dalla regione vicina e fornite come *input* al *layer* successivo della rete. L'operazione di convoluzione del bordo viene eseguita sull'*output* di ogni *layer* della rete. Nell'implementazione originale, al livello di *input* kNN è alimentato solo con coordinate di punti normalizzate, mentre nell'implementazione proposta si utilizzano tutte le funzionalità disponibili. Nello specifico, sono state aggiunte *features* della componente radiometrica, espresse come RGB o HSV, e vettori normali.

La Figura 4 mostra la struttura complessiva della rete. In *input* viene dato un blocco della scena (pari a un cubo con altezza "infinita"), composto da 12 *features* per ogni punto: le coordinate XYZ, le coordinate normalizzate X'Y'Z', la componente

radiometrica (canali HSV) e le normali. Questi blocchi passano attraverso 4 *layer EdgeConv* e un livello di *max pooling* per estrarre le *features* globali del blocco. Le coordinate XYZ originali vengono mantenute per tenere conto del posizionamento dei punti nell'intera scena, mentre le coordinate normalizzate rappresentano il posizionamento all'interno di ogni blocco. Il modulo kNN viene alimentato solo con coordinate normalizzate e sia le coordinate originali che quelle normalizzate vengono utilizzate come *features* di *input* per la rete neurale. I canali RGB sono stati convertiti in canali HSV in due passaggi: prima vengono normalizzati tra 0 e 1, quindi vengono convertiti in canali HSV utilizzando la funzione *rgb2hsv()* della libreria *scikit-image* implementata in *Python*. Questa conversione è utile perché i singoli canali H, S e V sono indipendenti l'uno dall'altro, ognuno di essi ha un'informazione di tipologia diversa, rendendoli delle *features* indipendenti. I canali R, G e B sono al contrario in qualche modo correlati tra loro, condividono infatti una parte dello stesso tipo di dati e quindi non si possono usare separatamente.

La scelta di utilizzare normali e HSV è dovuta a diversi motivi. Da un lato la componente RGB, basata sui sensori utilizzati nell'acquisizione dati, è il più delle volte presente come proprietà della nuvola di punti e quindi si è deciso di sfruttare appieno questo tipo di dato; dall'altro le componenti RGB definiscono le proprietà radiometriche della nuvola di punti, mentre le normali definiscono alcune proprietà geometriche. In questo modo si utilizzano come *input* alla rete neurale diversi tipi di informazioni. Inoltre, la decisione di convertire i dati RGB in HSV si basa su altri lavori di ricerca (Surol et al., 2002) che, anche se sviluppati per compiti diversi, dimostrano l'efficacia di questa operazione.

Il primo *layer EdgeConv* è stato modificato in modo che il kNN possa usare anche il colore e le normali per poter selezionare i *k*-vicini per ogni punto. Infine, attraverso 3 *layer* convoluzionali e un *layer dropout*, si produce lo stesso blocco di punti ma con punteggi sulla probabilità della segmentazione (uno per ogni classe da riconoscere). L'*output* della segmentazione sarà dato dalla classe con il punteggio più alto.

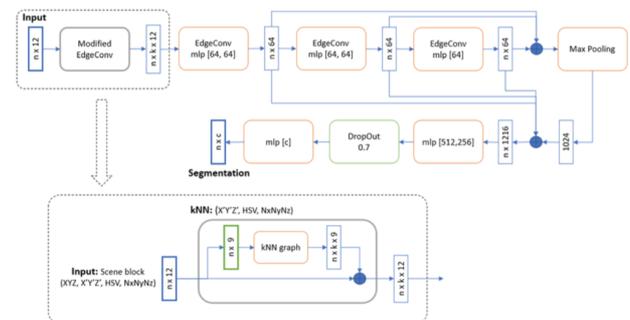


Figura 4. Illustrazione dell'architettura della DGCNN modificata.

4. RISULTATI

In questa sezione vengono riportati i risultati degli esperimenti. Oltre alle prestazioni della DGCNN modificata, sono presentate anche le prestazioni di PointNet, PointNet++, PCNN e DGCNN. Gli esperimenti sono divisi in due fasi:

- Nella prima (§ 4.1), le reti sono addestrate scegliendo gli iperparametri migliori per segmentare semanticamente il dataset. A tal fine è stata considerata un'unica scena simmetrica, di cui ne è stata utilizzata una parte annotata per l'addestramento della rete e la parte restante per la valutazione delle prestazioni. Questi primi esperimenti sono stati condotti sulla scena della Chiesa del Trompone

(TR_church) in quanto presenta un buon grado di simmetria, che permette di suddividerla in parti con caratteristiche simili e con quasi tutte le classi considerate (9 su 10). Tale impostazione risolve il problema di annotare automaticamente una scena che è stata annotata solo parzialmente manualmente. Sebbene ciò possa avere applicazioni pratiche e possa accelerare il processo di annotazione di un'intera scena, l'obiettivo è valutare l'annotazione automatica di una scena che non è mai stata vista prima dalla rete, pertanto si è resa necessaria una seconda fase.

- Nella seconda (§ 4.2), le reti sono addestrate con 10 scene diverse e quella rimanente viene predetta automaticamente nella fase di test.

La segmentazione dell'intera nuvola di punti in sotto parti (blocchi) è una fase di pre-elaborazione necessaria per tutte le architetture neurali analizzate. Per ogni blocco deve essere campionato un numero fisso di punti. Ciò è dovuto al fatto che le reti neurali richiedono un numero costante di punti come *input* e sarebbe inoltre impossibile, dal punto di vista computazionale, fornire alle reti tutti i punti della nuvola contemporaneamente.

4.1 Segmentazione della scena parzialmente annotata

In questa fase sono state valutate due diverse impostazioni: una *k-fold cross-validation* e una singola suddivisione del dataset annotato in set di addestramento e set di test. Nel primo caso il numero complessivo di campioni di prova è piccolo e la rete è addestrata su più campioni. Nel secondo caso, lo stesso numero di campioni viene utilizzato per addestrare e valutare la rete, portando possibilmente a risultati molto diversi. Per completezza, sono state testate entrambe le impostazioni.

Nella prima impostazione, la scena TR_church è stata divisa in 6 parti ed è stata eseguita una *cross-validation* con un valore di *k* pari a 6, come mostrato in Figura 5. Sono state testate diverse combinazioni di iperparametri delle varie reti per poter verificare quale fosse la migliore. Si veda la Tabella 1, dove l'accuratezza media è derivata dal calcolo dell'accuratezza di ciascun test (*fold*), quindi dalla media.

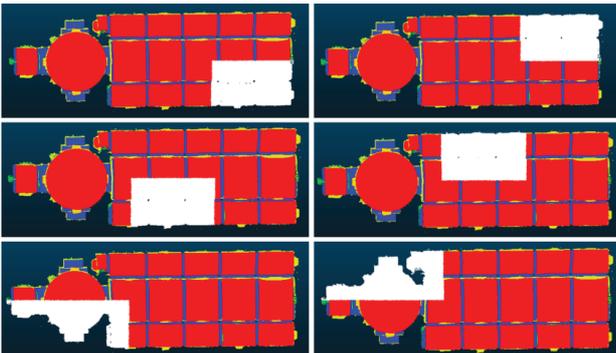


Figura 5. 6-fold Cross Validation sulla scena TR_church. La *fold* bianca in ogni esperimento è la parte utilizzata per il test

Rete	Features	Acc. media
PointNet	XYZ	0.543
PointNet++	XYZ	0.459
PCNN	XYZ	0.742
DGCNN	XYZ+RGB	0.897
DGCNN-Mod-1	XYZ+Normali	0.781
DGCNN-Mod-2	XYZ+HSV+Normali	0.918

Tabella 1. 6-fold Cross-Validation sulla scena Trompone. Sono state scelte diverse combinazioni di iperparametri per le varie reti dello stato dell'arte.

Per quanto riguarda le fasi di pre-elaborazione della nuvola di punti, che consiste nel segmentare l'intera scena in blocchi e, per ogni blocco, campionare un numero di punti, sono state scelte, per ogni modello valutato, le impostazioni di default. I blocchi di PointNet e PointNet++ sono di dimensioni 2×2 metri e 4096 punti per il campionamento delle nuvole. Nel caso della DGCNN, sono stati utilizzati blocchi di dimensione 1×1 metri e 4096 punti per blocco. Infine, la rete PCNN è stata testata utilizzando lo stesso campionamento della DGCNN (1×1), ma utilizzando 2048 punti, poiché questa è l'impostazione predefinita utilizzata nella PCNN.

Si è inoltre testata anche la PCNN fornendo 4096 punti per blocco, ma i risultati sono stati leggermente peggiori. Si è inoltre notato che le prestazioni migliorano leggermente utilizzando le *features* del colore rappresentate come mappa colori HSV. La rappresentazione HSV (tonalità, saturazione, valore) è nota per essere più strettamente allineata con la percezione umana dei colori e, rappresentando i colori come tre variabili indipendenti, consente di tenere conto delle variazioni, dovute alle ombre e alle diverse condizioni di luce.

Nella seconda impostazione degli esperimenti la scena della Chiesa del Trompone è stata suddivisa a metà lungo l'asse di simmetria, scegliendo il lato sinistro per l'addestramento e il lato destro per il test. Il lato sinistro è stato ulteriormente suddiviso in un set di addestramento (80%) e un set di validazione (20%). Tale set di validazione è stato utilizzato per testare l'accuratezza complessiva alla fine di ogni epoca di addestramento e la valutazione delle *performance* è stata eseguita sulla parte rimanente di test (lato destro). Nella Tabella 2 sono riportate le metriche delle reti. I risultati ripostati sono quelli ottenuti con le migliori combinazioni di iperparametri ottenute dall'esperimento di convalida incrociata.

Rete	Acc.	Prec	Rec	F ₁ score
PointNet	0.307	0.405	0.306	0.287
PointNet++	0.441	0.480	0.487	0.448
PCNN	0.623	0.642	0.608	0.636
DGCNN	0.733	0.721	0.733	0.707
DGCNN-Mod-2	0.743	0.748	0.742	0.722

Tabella 2. La scena è stata divisa in 3 parti: *Training*, *Validation*, *Test*. Media delle metriche calcolate sulle diverse parti: accuratezza per Training e Test; precisione, recall, F₁-score e support per il test.

La Figura 6 mostra la scena del test annotata manualmente (*ground truth*) e i risultati della segmentazione automatica ottenuti con l'approccio proposto.

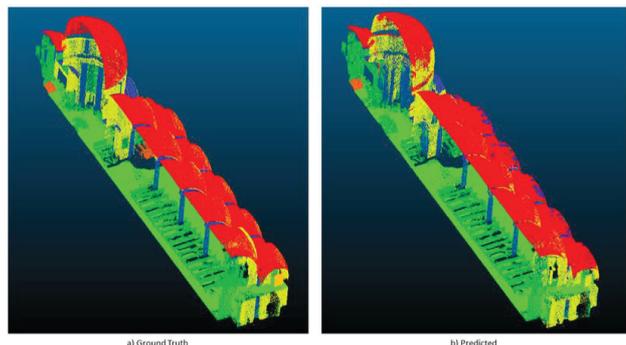


Figura 6. *Ground Truth* e nuvola di punti predetta, utilizzando la DGCNN modificata, sul lato di test di Trompone.

4.2 Segmentazione di una scena mai vista dalla rete neurale

Nella seconda fase sperimentale sono state utilizzate tutte le scene a disposizione: 9 per l'addestramento, 1 per la validazione (Ghiffa) e 1 per il test (SMV). Come nella fase precedente, sono state valutate le reti dello stato dell'arte, confrontando i risultati con l'approccio basato sulla DGCNN modificata. Nella Tabella 3 sono riportate le prestazioni complessive per ogni modello testato. La Figura 7 mostra la matrice di confusione della segmentazione dell'ultimo esperimento: 9 scene per l'addestramento, 1 scena per la validazione e 1 scena per il test. Il miglioramento delle prestazioni è più evidente rispetto agli esperimenti precedenti, portando a un miglioramento di circa 0,8 nella precisione complessiva e nel punteggio F1score. Anche l'Intersection over Union (IoU) aumenta. Per alcune classi i valori di precisione e recall sono inferiori alla DGCNN originale. Tuttavia, questo approccio alla DGCNN generalmente migliora le prestazioni in termini di punteggio F1score.

Rete	Test	Prec.	Rec.	F1score
PointNet	0.351	0.536	0.351	0.269
PointNet++	0.528	0.532	0.528	0.479
PCNN	0.629	0.653	0.622	0.635
DGCNN	0.740	0.768	0.740	0.738
DGCNN-Mod-2	0.825	0.809	0.825	0.814

Tabella 3. Risultati dei test effettuati addestrando la rete con 9 nuvole di punti e testandola su una scena sconosciuta.

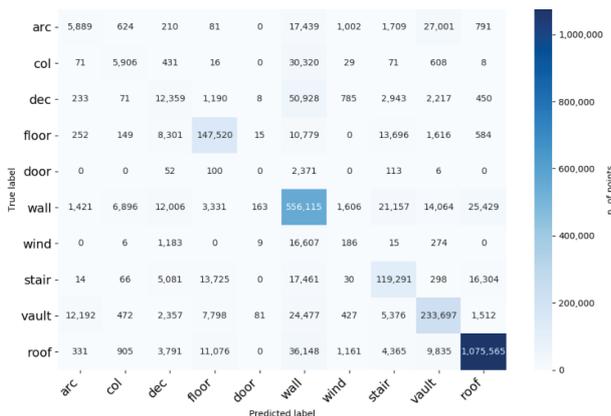


Figura 7. Matrice di confusione per l'ultimo esperimento: 9 scene per l'addestramento, 1 scena per la validazione e 1 scena per il test.

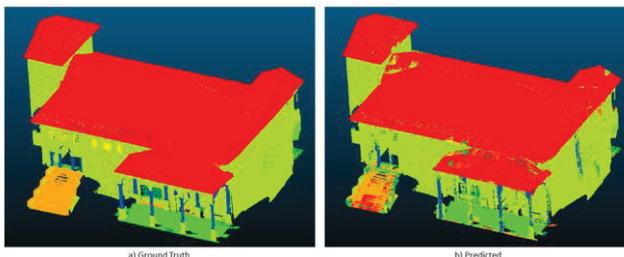


Figura 8. Ground Truth e nuvola di punti predetta, utilizzando il la DGCNN modificata.

5. DISCUSSIONI

In questa sezione si considerano alcuni aspetti della ricerca (e le sfide) che vale la pena approfondire. Prima di tutto, guardando il

primo assetto sperimentale, le prestazioni sono peggiori di quelle ottenute nell'esperimento *k-fold* (facendo riferimento alla Tabella 2). Questo è probabilmente dovuto al fatto che la rete ha meno punti su cui apprendere. I risultati in fase di test ottenuti con i criteri proposti confermano che, considerando HSV+Normali, la rete riesce ad apprendere *features* di livello superiore delle diverse classi. Inoltre, osservando la Figura 6, si può notare che l'utilizzo delle impostazioni qui descritte aiuta a migliorare la predizione della classe delle volte, aumentandone la precisione, *recall* e l'IoU, così come le colonne e le scale.

Considerando la seconda impostazione sperimentale, si può osservare che tutti gli approcci non riescono a riconoscere le classi con basso numero di punti, come porte, finestre e archi. Inoltre, per queste classi si osserva un'elevata variabilità nelle forme e geometrie all'interno del dataset, questo probabilmente contribuisce alla bassa precisione ottenuta dalle reti.

Ulteriori considerazioni possono essere fatte valutando la matrice di confusione mostrata nella Figura 7. Essa rivela, ad esempio, che gli archi sono spesso confusi con le volte, poiché condividono chiaramente le caratteristiche geometriche, mentre le colonne sono spesso confuse con i muri. Quest'ultimo comportamento può essere dovuto alla presenza di lesene, che sono annotate come colonne, ma hanno una forma simile ai muri. La natura sbilanciata del numero di punti per classe è chiaramente evidenziata nella Figura 9.

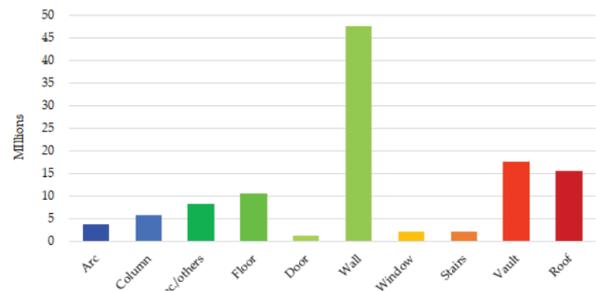


Figura 9. Numero di punti per classe

Inoltre, se si analizzano le classi singolarmente, è possibile notare che i valori più bassi sono in *Arc*, *Dec*, *Door* e *Window*. Più in dettaglio:

- **Arco:** la geometria degli elementi di questa classe è molto simile a quella delle volte e, sebbene le dimensioni degli archi non siano simili a queste ultime, il più delle volte sono attigui alle volte stesse, quasi una continuazione di questi elementi. Per tali motivi il risultato è in parte giustificabile e ha portato alla fusione di queste due classi (Matrone et al. 2020a, Matrone et al. 2020b).
- **Decorazioni:** in questa classe, che può essere definita anche "altri" o "non assegnati", sono inclusi tutti gli elementi che non fanno parte delle altre classi (come panchine, quadri, confessionali ...). Pertanto, non è da considerare appieno tra i risultati.
- **Porta:** il risultato nullo è quasi certamente dovuto al numero molto basso di punti presenti in questa classe (Figura 9). Infatti, nei casi studio proposti di CH, è più comune trovare grandi archi che segnano il passaggio da uno spazio all'altro e le porte sono appena presenti. Inoltre, durante le fasi di rilievo, spesso le porte risultano aperte o con occlusioni, generando una visione e un'acquisizione parziale di questi elementi.
- **Finestra:** in questo caso il risultato non è dovuto al basso numero di finestre presenti nel caso studio, ma all'elevata eterogeneità tra loro. Infatti, sebbene il numero di punti in questa classe sia maggiore, le forme delle aperture sono

molto diverse tra loro. Inoltre, essendo per lo più composte da superfici vetrate, queste superfici non vengono rilevate dai sensori coinvolti come il TLS, quindi, a differenza dell'uso delle immagini, in questo caso il numero di punti utili a descrivere questi elementi è ridotto.

6. CONCLUSIONI

La segmentazione semantica delle nuvole di punti è un compito rilevante nell'ambito del patrimonio culturale digitale in quanto consente di riconoscere automaticamente diversi tipi di elementi architettonici e storici, risparmiando tempo e velocizzando il processo di analisi delle nuvole di punti acquisite e di modellazione 3D. Nel contesto degli edifici storici, la segmentazione semantica della nuvola di punti è resa particolarmente impegnativa dalla complessità e dall'elevata variabilità degli oggetti da rilevare. In questo lavoro, è stata fornita una prima valutazione delle tecniche di segmentazione delle nuvole di punti basate su metodi di DL, a partire dallo stato dell'arte nel contesto degli edifici storici. Oltre a confrontare le prestazioni degli approcci esistenti, è qui proposta una modifica della rete che ne aumenta l'accuratezza della segmentazione, dimostrando l'efficacia e l'idoneità del metodo qui proposto. Tale architettura è basata su una versione modificata della DGCNN ed è stata testata su parte di un dataset appositamente creato: l'ArCH (*Architectural Cultural Heritage*) dataset. I risultati dimostrano che la metodologia proposta è adatta alla segmentazione semantica di nuvole di punti con applicazioni pertinenti. La ricerca parte dall'idea di raccogliere un dataset DCH da condividere con la comunità scientifica insieme ai codici sorgente del *framework*, per poter confrontare il metodo proposto, e permetterne modifiche e ottimizzazioni. L'articolo descrive uno dei test più estesi basati sui dati DCH e ha un enorme potenziale nel campo dell'HBIM, al fine di rendere accessibile e più veloce il processo *scan-to-BIM*.

Tuttavia, i risultati ottenuti hanno evidenziato alcune lacune e sfide aperte che è giusto menzionare. Innanzitutto, il *framework* non è in grado di valutare le prestazioni di accuratezza rispetto alle tecniche di acquisizione. In altre parole, si cercherà di scoprire, con test futuri, se l'adozione di nuvole di punti acquisite con altri metodi cambia le prestazioni della metodologia proposta. Inoltre, la dimensione dei punti per classi è sbilanciata e non omogenea, come dimostra anche la matrice di confusione. Questo collo di bottiglia può essere risolto annotando un dataset più dettagliato o creando nuvole di punti sintetiche. Il gruppo di ricerca sta concentrando i suoi sforzi anche in questa direzione (Pierdicca et al., 2019). Infine, nei lavori futuri, si cercherà di migliorare e integrare meglio il *framework* con architetture più efficaci, al fine di migliorarne le prestazioni, oltre ad approfondire studi e comparazioni tra l'approccio ML (Grilli et al., 2019) e DL. I primi risultati in tal senso (Matrone et al., 2020b) mostrano come l'aggiunta di *features* 3D, basate sugli *eigenvalue*, aumentino ulteriormente le prestazioni della rete e aiutino nel riconoscimento di quelle classi con un minor numero di punti, tuttavia rimangono ancora alcuni aspetti da migliorare quali l'alta richiesta di potenza computazionale e i tempi di addestramento se basati su nuovi set di dati.

BIBLIOGRAFIA

Armeni, I., Sener, O., Zamir, A.R., Jiang, H., Brilakis, I., Fischer, M. and Savarese, S., 2016. 3d semantic parsing of large-scale indoor spaces. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1534-1543).

Atzmon, M., Maron, H. and Lipman, Y., 2018. Point convolutional neural networks by extension operators. *arXiv preprint arXiv:1803.10091*.

Balletti, C., D'agnano, F., Guerra, F. and Vernier, P., 2016. From point cloud to digital fabrication: A tangible reconstruction of Ca' Venier dei Leoni, the Guggenheim Museum in Venice. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 3, p.43.

Barazzetti, L., Banfi, F., Brumana, R., Oreni, D., Previtali, M. and Roncoroni, F., 2015. HBIM and augmented information: towards a wider user community of image and range-based reconstructions. In *25th International CIPA Symposium 2015* (Vol. 40, pp. 35-42).

Barazzetti, L. and Previtali, M., 2019. Vault Modeling with Neural Networks. In *8th International Workshop on 3D Virtual Reconstruction and Visualization of Complex Architectures, 3D-ARCH 2019* (Vol. 42, No. 2, pp. 81-86). Copernicus GmbH.

Barsanti, S.G., Guidi, G. and De Luca, L., 2017. Segmentation of 3D models for cultural heritage structural analysis—some critical issues. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 4, p.115.

Bitelli, G., Dellapasqua, M., Girelli, V.A., Sanchini, E. and Tini, M.A., 2017. 3D Geomatics techniques for an integrated approach to cultural heritage knowledge: the case of san michele in acerboli's church in santarcangelo di romagna. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, 42.

Bolognesi, C. and Garagnani, S., 2018. From a point cloud survey to a mass 3D modelling: Renaissance HBIM in Poggio a Caiano.

Borin, P. and Cavazzini, F., 2019. Condition assessment of rc bridges. integrating machine learning, photogrammetry and bim. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*.

Bronzino, G.P.C., Grasso, N., Matrone, F., Osello, A. and Piras, M., 2019. Laser-visual-inertial odometry based solution for 3D heritage modeling: the sanctuary of the blessed virgin of Trompone. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*.

Bruno, N. and Roncella, R., 2018. A restoration oriented HBIM system for cultural heritage documentation: the case study of Parma Cathedral. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, 42(2).

Capone, M. and Lanzara, E., 2019. Scan-to-BIM vs 3d ideal model HBIM: parametric tools to study domes geometry. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*.

Fregonese, L., Taffurelli, L., Adami, A., Chiarini, S., Cremonesi, S., Helder, J. and Spezzoni, A., 2017. Survey and modelling for the BIM of Basilica of San Marco in Venice. In *2017 TC II and CIPA-3D Virtual Reconstruction and Visualization of Complex Architectures* (Vol. 42, No. 2W3, pp. 303-310). International Society for Photogrammetry and Remote Sensing.

Geiger, A., Lenz, P., Stiller, C. and Urtasun, R., 2013. Vision meets robotics: The kitti dataset. *The International Journal of Robotics Research*, 32(11), pp.1231-1237.

Grilli, E., Dinunno, D., Petrucci, G. and Remondino, F., 2018. From 2D to 3D supervised segmentation and classification for cultural heritage applications. In *ISPRS TC II Mid-term Symposium "Towards Photogrammetry 2020"*, 42, 42, pp. 399-406.

Grilli, E., Özdemir, E. and Remondino, F., 2019. Application of machine and deep learning strategies for the classification of heritage point clouds. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*.

Grilli, E. and Remondino, F., 2019. Classification of 3D digital heritage. *Remote Sensing*, 11(7), p.847.

- Hackel, T., Savinov, N., Ladicky, L., Wegner, J.D., Schindler, K. and Pollefeys, M., 2017. Semantic3d. net: A new large-scale point cloud classification benchmark. *arXiv preprint arXiv:1704.03847*.
- Llamas, J., M Lerones, P., Medina, R., Zalama, E. and Gómez-García-Bermejo, J., 2017. Classification of architectural heritage images using deep learning techniques. *Applied Sciences*, 7(10), p.992.
- Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G. and Johnson, B.A., 2019. Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS journal of photogrammetry and remote sensing*, 152, pp.166-177.
- Macher, H., Landes, T. and Grussenmeyer, P., 2017. From point clouds to building information models: 3D semi-automatic reconstruction of indoors of existing buildings. *Applied Sciences*, 7(10), p.1030.
- Masiero, A., Fissore, F., Guarnieri, A., Pirotti, F., Visintini, D. and Vettore, A., 2018. Performance evaluation of two indoor mapping systems: Low-cost UWB-aided photogrammetry and backpack laser scanning. *Applied Sciences*, 8(3), p.416.
- Mathias, M., Martinovic, A., Weissenberg, J., Haegler, S. and Van Gool, L., 2011. Automatic architectural style recognition. *ISPRAR*, 3816, pp.171-176.
- Matrone, F., Lingua, A., Pierdicca, R., Malinverni, E. S., Paolanti, M., Grilli, E., Remondino, F., Murtiyoso, A., and Landes, T., 2020a. A Benchmark For Large-Scale Heritage Point Cloud Semantic Segmentation. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.*, XLIII-B2-2020, pp. 1419–1426.
- Matrone, F., Grilli, E., Martini, M., Paolanti, M., Pierdicca, R., Remondino, F., 2020b. Comparing Machine and Deep Learning Methods for Large 3D Heritage Semantic Segmentation. *ISPRS Int. J. Geo-Inf.*, 9, 535.
- Munoz, D., Bagnell, J.A., Vandapel, N. and Hebert, M., 2009, June. Contextual classification with functional max-margin markov networks. In *2009 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 975-982). IEEE.
- Murtiyoso, A. and Grussenmeyer, P., 2019. Automatic heritage building point cloud segmentation and classification using geometrical rules. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*.
- Oreni, D., Brumana, R., Della Torre, S. and Banfi, F., 2017. Survey, HBIM and conservation plan of a monumental building damaged by earthquake.
- Osello, A., Lucibello, G. and Morgagni, F., 2018. HBIM and virtual tools: A new chance to preserve architectural heritage. *Buildings*, 8(1), p.12.
- Oses, N., Dornaika, F. and Moujahid, A., 2014. Image-based delineation and classification of built heritage masonry. *Remote Sensing*, 6(3), pp.1863-1889.
- Pierdicca, R., Mameli, M., Malinverni, E.S., Paolanti, M. and Frontoni, E., 2019, June. Automatic Generation of Point Cloud Synthetic Dataset for Historical Building Representation. In *International Conference on Augmented Reality, Virtual Reality and Computer Graphics* (pp. 203-219). Springer, Cham.
- Qi, C.R., Su, H., Mo, K. and Guibas, L.J., 2017a. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 652-660).
- Qi, C.R., Yi, L., Su, H. and Guibas, L.J., 2017b. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. In *Advances in neural information processing systems* (pp. 5099-5108).
- Quattrini, R., Pierdicca, R. and Morbidoni, C., 2017. Knowledge-based data enrichment for HBIM: Exploring high-quality models using the semantic-web. *Journal of Cultural Heritage*, 28, pp.129-139.
- Song, S. and Xiao, J., 2016. Deep sliding shapes for a modal 3d object detection in rgb-d images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 808-816).
- Spina, S., Debattista, K., Bugeja, K. and Chalmers, A., 2011, October. Point cloud segmentation for cultural heritage sites. In *Proceedings of the 12th International conference on Virtual Reality, Archaeology and Cultural Heritage* (pp. 41-48).
- Stathopoulou, E.K. and Remondino, F., 2019. Semantic photogrammetry: boosting image-based 3D reconstruction with semantic labeling. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.*, 42(2), p.W9.
- Sural, S., Qian, G. and Pramanik, S., 2002, September. Segmentation and histogram generation using the HSV color space for image retrieval. In *Proceedings. International Conference on Image Processing* (Vol. 2, pp. II-II). IEEE.
- Tamke, M., Evers, H.L., Zwierzycki, M., Wessel, R., Ochmann, S., Vock, R. and Klein, R., 2016, September. An automated approach to the generation of structured building information models from unstructured 3D point cloud scans. In *Proceedings of IASS Annual Symposia* (Vol. 2016, No. 17, pp. 1-10). International Association for Shell and Spatial Structures (IASS).
- Tang, P., Huber, D., Akinci, B., Lipman, R. and Lytle, A., 2010. Automatic reconstruction of as-built building information models from laser-scanned point clouds: A review of related techniques. *Automation in construction*, 19(7), pp.829-843.
- Thomson, C. and Boehm, J., 2015. Automatic geometry generation from point clouds for BIM. *Remote Sensing*, 7(9), pp.11753-11775.
- Wang, W., Yu, R., Huang, Q. and Neumann, U., 2018. Sgpn: Similarity group proposal network for 3d point cloud instance segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 2569-2578).
- Wang, Y., Sun, Y., Liu, Z., Sarma, S.E., Bronstein, M.M. and Solomon, J.M., 2019. Dynamic graph cnn for learning on point clouds. *Acm Transactions On Graphics (tog)*, 38(5), pp.1-12.
- Weinmann, M., Jutzi, B., Hinz, S. and Mallet, C., 2015. Semantic point cloud interpretation based on optimal neighborhoods, relevant features and efficient classifiers. *ISPRS Journal of Photogrammetry and Remote Sensing*, 105, pp.286-304.
- Wu, Z., Song, S., Khosla, A., Yu, F., Zhang, L., Tang, X. and Xiao, J., 2015. 3d shapenets: A deep representation for volumetric shapes. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1912-1920).
- Zaheer, M., Kottur, S., Ravanbakhsh, S., Poczos, B., Salakhutdinov, R.R. and Smola, A.J., 2017. Deep sets. In *Advances in neural information processing systems* (pp. 3391-3401).
- Zhang, K., Hao, M., Wang, J., de Silva, C.W. and Fu, C., 2019. Linked dynamic graph CNN: Learning on point cloud via linking hierarchical features. *arXiv preprint arXiv:1904.10014*.

UN BENCHMARK PER LA SEGMENTAZIONE SEMANTICA DI NUVOLE DI PUNTI DI BENI CULTURALI

A BENCHMARK FOR LARGE-SCALE HERITAGE POINT CLOUD SEMANTIC SEGMENTATION

F. Matrone¹, A. Lingua^{1,2}, R. Pierdicca³, E.S. Malinverni³, M. Paolanti⁴, E. Grilli⁵, F. Remondino⁵, A. Murtiyoso⁶, T. Landes⁶

¹ DIATI, Politecnico di Torino, Torino, Italy - Email: <francesca.matrone><andrea.lingua>@polito.it

² PIC4SeR, Politecnico Interdepartmental Center for Service Robotics, Politecnico di Torino, Torino, Italy

³ DICEA, Università Politecnica delle Marche, Ancona, Italy - Email: <r.pierdicca><e.s.malinverni>@staff.univpm.it

⁴ DII, Università Politecnica delle Marche, Ancona, Italy - Email: m.paolanti@staff.univpm.it

⁵ 3D Optical Metrology (3DOM) unit, Bruno Kessler Foundation (FBK), Trento, Italy – Email: <grilli><remondino>@fbk.eu

⁶ Photogrammetry and Geomatics Group, ICube Laboratory UMR 7357, INSA Strasbourg, France – <arnadi.murtiyoso><tania.landes>@insa-strasbourg.fr

PAROLE CHIAVE: Benchmark, beni culturali, nuvole di punti, segmentazione semantica, classificazione, machine learning, deep learning

KEYWORDS: Benchmark, 3D heritage, point cloud, semantic segmentation, classification, machine learning, deep learning

RIASSUNTO

La mancanza di dati di benchmark per la segmentazione semantica di nuvole di punti dei beni culturali sta ostacolando lo sviluppo di soluzioni di classificazione automatica in questo campo. I dati 3D e le nuvole di punti del nostro patrimonio culturale rappresentano strutture geometriche complesse con classi non convenzionali, le quali impediscono la semplice implementazione dei metodi già disponibili, sviluppati in altri campi o per altri tipi di dati. La segmentazione semantica dei dati 3D del patrimonio aiuterebbe la comunità nella migliore comprensione e analisi dei gemelli digitali (*digital twins*), faciliterebbe le operazioni di salvaguardia e supporterebbe molte altre attività legate al settore dei beni culturali. In questo contributo si presenta il primo benchmark con milioni di punti 3D annotati manualmente appartenenti a scenari del patrimonio, realizzati per facilitare lo sviluppo, l'addestramento, il test e la valutazione di metodi e algoritmi di apprendimento automatico nel campo dei beni architettonici. Il benchmark proposto, disponibile su <http://archdataset.polito.it/>, comprende set di dati e risultati di classificazione finalizzati a migliorare i confronti e approfondire i punti di forza e di debolezza dei diversi approcci di *machine* e *deep learning* per la segmentazione semantica della nuvola di punti del patrimonio, oltre a promuovere una forma di crowdsourcing per arricchire il database già annotato.

ABSTRACT

The lack of benchmarking data for the semantic segmentation of digital heritage scenarios is hampering the development of automatic classification solutions in this field. Heritage 3D data feature complex structures and uncommon classes that prevent the simple deployment of available methods developed in other fields and for other types of data. The semantic classification of heritage 3D data would support the community in better understanding and analysing digital twins, facilitate restoration and conservation work, etc. This paper presents the first benchmark with millions of manually labelled 3D points belonging to heritage scenarios, realised to facilitate the development, training, testing and evaluation of machine and deep learning methods and algorithms in the heritage field. The proposed benchmark, available at <http://archdataset.polito.it/>, comprises datasets and classification results for better comparisons and insights into the strengths and weaknesses of different machine and deep learning approaches for heritage point cloud semantic segmentation, in addition to promoting a form of crowdsourcing to enrich the already annotated database.

1. INTRODUZIONE

La crescente facilità di acquisizione delle nuvole di punti, soprattutto grazie agli sviluppi di soluzioni automatizzate basate su immagini, metodi SLAM (*simultaneous localization and mapping*) e sistemi di scansione laser, ha creato un crescente interesse nella comunità scientifica verso l'uso, l'interpretazione e lo sfruttamento diretto delle nuvole di punti per molti scopi. Parallelamente, nel campo dei Beni Culturali (BBCC), l'HBIM (*Historical Building Information Modeling*) ha ottenuto una particolare attenzione da parte degli esperti, poiché consente di gestire i dati del patrimonio architettonico, sia in modo geometrico che informativo (Bruno e Roncella, 2018). Come è noto, se le nuvole di punti forniscono un punto di partenza per

tale modellazione, il processo *scan-to-BIM* è ancora affidato a operazioni completamente manuali; gli esperti devono gestire set di dati ampi e complessi, senza l'ausilio di alcun metodo automatico o semiautomatico per riconoscere e rimodellare gli elementi 3D (Bitelli et al., 2017). Ovviamente questo processo richiede molto tempo e porta allo spreco di informazioni, vista l'inevitabile semplificazione esercitata.

In questo scenario, il ritorno del *Deep Learning* (DL) in diversi campi di ricerca è stato di notevole importanza (Griffiths e Boehm, 2019). Le reti neurali, in particolare le *Deep Neural Network* (DNN), si sono affermate come la tecnologia più efficiente per le attività basate sull'apprendimento artificiale (Paolanti et al., 2019; Bello et al., 2020). Tuttavia, nonostante le DNN si siano rivelate molto promettenti per la gestione e il

riconoscimento dei dati 3D (Wang et al., 2019), per i BBCC le operazioni manuali sembrano più affidabili, almeno per identificare gli elementi architettonici da nuvole di punti (Murtiyoso e Grussenmeyer, 2019). Ci sono infatti molte ragioni per questo scetticismo: innanzitutto, i BBCC hanno geometrie complesse, che possono essere descritte solo con un alto livello di dettaglio; inoltre, le forme irregolari unite all'unicità degli oggetti, rendono ardue le tecniche di apprendimento supervisionato per i dati 3D.

Oltre alla complessità intrinseca dei dati 3D, soprattutto se confrontati con quelli 2D (es. immagini), ci sono altre limitazioni che ostacolano lo sfruttamento delle DNN per i BBCC; da una parte la mancanza di dati di addestramento, dall'altra l'alta potenza computazionale richiesta. Sebbene quest'ultima sarà risolta dai continui progressi tecnologici, si è ancora lontani dal consentire a un sistema di imparare da un set di dati sui BBCC annotato e di generalizzare su scene sconosciute alla rete. L'annotazione manuale è però costosa e richiede tempo (anche se più affidabile), ed esiste una sorta di reticenza a condividere i dati 3D con la comunità scientifica.

Con lo scopo principale di investire molti più sforzi su queste linee di ricerca, gli autori forniscono un ampio dataset di architetture dei BBCC, che aspira a diventare il benchmark di riferimento nel campo. È il primo set di dati di nuvole di punti specificamente rilasciato per l'ambito dei BBCC, che comprende dati raccolti con rilievi da laser scanner e fotogrammetria, fornendone anche l'annotazione semantica (*ground truth*).

Questo contributo mira a presentare un nuovo set di dati di riferimento per la classificazione delle nuvole di punti 3D (denominato ArCH dataset – *Architectural Cultural Heritage*) con milioni di punti annotati manualmente appartenenti a scenari del patrimonio. Il benchmark realizzato nasce dalla collaborazione di diverse Università e istituti di ricerca (Politecnico di Torino, Università Politecnica delle Marche, FBK e l'INSA - *Institut National des Sciences Appliquées* - di Strasburgo). È un dataset unico in quanto offre, per la prima volta alla comunità scientifica, nuvole di punti annotate che descrivono scene del patrimonio. Queste nuvole di punti, annotate con 10 classi, hanno lo scopo di facilitare lo sviluppo, l'addestramento, il test e la valutazione di algoritmi di apprendimento automatico (*Machine learning - ML*), nonché il suo sottoinsieme di metodi di apprendimento profondo (*Deep learning - DL*) nel campo del patrimonio. Per un utilizzo più proficuo di questo benchmark, oltre al *download* gratuito di tutti i dati, sono pubblicati sul sito (www.archdataset.polito.it) anche i risultati degli approcci testati, fornendo classifiche su quelli più performanti.

2. LAVORI PRECEDENTI

Nella comunità geomatica sono stati proposti diversi benchmark. I dati 3D annotati consentono agli utenti di testare e validare i propri algoritmi, oltre a migliorare le fasi di addestramento per gli approcci di apprendimento sia automatico che profondo. Tra i benchmark esistenti, vale la pena citare ModelNet 40 (Wu et al., 2015) con più di 100 mila modelli CAD di oggetti, principalmente mobili, di 40 diverse categorie; KITTI (Geiger et al., 2013) che include immagini e scansioni laser per la navigazione autonoma; il dataset Sydney Urban Objects (De Deuge et al., 2013) acquisito in ambienti urbani con 26 classi e 631 scansioni; Semantic3D (Hackel et al., 2017) con scene urbane come chiese, strade, ferrovie e piazze; S3DIS (Armeni et al., 2016) che include principalmente uffici e il set di dati Oakland 3-D Point Cloud (Munoz et al., 2009) costituito da nuvole di punti acquisite da *Mobile Mapping Systems* (MMS) in ambiente urbano. Inoltre, vale la pena menzionare altri set di dati specifici, come iQmulus (Vallet et al., 2015), The Cityscapes Dataset

(Cordts et al., 2016), Paris-rue-Madame (Serna et al., 2014), Paris- Lille-3D (Roynard et al., 2018), 3DOMcity (Özdemir et al., 2019) e MiMAP (Wang et al., 2018) per l'estrazione di *feature* per il BIM.

La maggior parte di questi set di dati raccoglie dati da ambienti urbani con nuvole di punti composte da circa 100.000 punti.

In questi scenari, le classi di oggetti sono abbastanza generali e quasi standard (ad es. terreno, strade, veicoli, vegetazione, edifici ecc.). Tuttavia, nel campo del patrimonio culturale, l'identificazione di categorie precise e puntuali è molto più complicata. Forma e colore non sono sempre legati a una specifica classe semantica e gli oggetti appartenenti alla stessa classe potrebbero avere forme completamente diverse, oltre a geometrie complesse. Inoltre, ad oggi, non ci sono ancora dataset pubblicati specifici sui BBCC immobili, con un adeguato livello di dettaglio.

Fino ad ora, la maggior parte dei set di dati disponibili del patrimonio architettonico annotato è costituito da immagini 2D, come il dataset Facades di Ecole Centrale Paris (ECP) (Teboul et al., 2010), eTRIMS (Korc e Forstner, 2009) e CMP Facade Database (Tyleček e Šára, 2013), che presentano tutti set di dati di immagini di facciate annotate manualmente da diverse città in tutto il mondo e diversi lessici architettonici.

Nel mondo del 2D, c'è il lavoro condotto da Llamas et al. (2017), dove per la prima volta sono state applicate le reti neurali convoluzionali (CNN) agli scenari del patrimonio. Gli autori hanno anche rilasciato un set di dati con più di 10.000 immagini con categorie quali altare, abside, campanile, colonna, cupola (interna ed esterna), contrafforte, *gargoyle*, vetrate e volte.

In questo contesto, diversi ricercatori hanno iniziato ad affrontare il tema della segmentazione semantica delle nuvole di punti dei BBCC all'interno del *framework* del ML e DL (Grilli et al., 2019a; Kharroubi et al., 2019; Murtiyoso e Grussenmeyer, 2020; Pierdicca et al., 2020). Tuttavia, la mancanza di un set di dati del patrimonio 3D appropriato non consente un confronto efficace tra metodi e risultati.

Proprio per questo, si propone l'ArCH dataset, in grado di stimolare la comunità scientifica su queste interessanti problematiche.

3. IL DATASET

Il dataset è composto da 17 nuvole di punti annotate e altre 10 non annotate. Quest'ultime potrebbero essere annotate dagli utenti e aggiunte, successivamente, al set di dati principale.

Molte delle scene incluse in questo benchmark fanno parte (o sono candidate) della Lista del Patrimonio Mondiale dell'UNESCO (WHL):

- la cappella della Cattedrale di Strasburgo all'interno della Grande Île, iscritta nel 1988;
- la sala aulica del Castello del Valentino (VAL) inclusa nelle Residenze Sabaude dal 1997;
- i Sacri Monti di Varallo (SMV) e Ghiffa (SMG) parte del sistema più ampio dei Sacri Monti del Piemonte e della Lombardia dal 2003;
- la chiesa di St. Pierre situata all'interno del *Neustadt* di Strasburgo iscritto nel 2017;
- i portici di Bologna candidati nel 2020.

Altre scene fanno comunque parte del patrimonio storico costruito e rappresentano vari periodi storici e lessici architettonici. Questa differenza potrebbe costituire un inconveniente nella definizione delle classi del dataset, in quanto introduce elementi di disomogeneità all'interno delle stesse classi. Tuttavia, dotare la rete neurale di elementi differenti

migliora la sua capacità di generalizzare tra vari casi di studio dei BBCC.

Tra le nuvole di punti (definite anche come “scene”) annotate del benchmark, 15 sono disponibili per l’addestramento degli algoritmi e 2 per le fasi di test. Comprendono chiese, cappelle, portici, logge, padiglioni e chiostri. Le 2 scene per il test (denominate A e B) hanno caratteristiche diverse:

- la prima (*A_SMG_portico*) rappresenta un edificio semplice, quasi simmetrico su un unico livello e con elementi geometrici più standard e ripetitivi (Figura 1);



Figura 1. Nuvola di punti del portico del Sacro Monte di Ghiffa (SMG).

- la seconda (*B_SMV_chapel_27to35*) rappresenta un edificio complesso, non simmetrico, strutturato su due livelli, rilevato sia all’interno che all’esterno, con diverse tipologie di volte, scale e finestre (Figura 2).

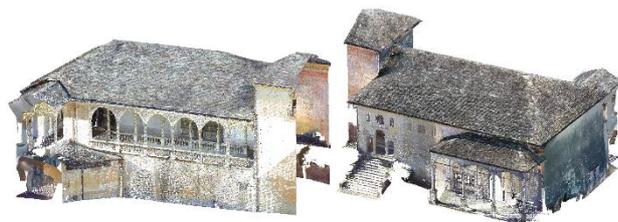


Figura 2. Nuvola di punti della seconda scena di test, che include le cappelle dalla 27 alla 35 del Sacro Monte of Varallo (prospetto nord e sud).

Queste due scene di test sono state scelte per (i) semplificare i confronti dei risultati, (ii) valutare l’efficacia degli algoritmi proposti e (iii) cercare di evidenziare la generalizzazione e la capacità di apprendimento delle reti non solo su una scena relativamente semplice ma anche su una più complessa.

3.1 Acquisizione dati

I dati 3D che compongono il benchmark (Tabella 1) sono stimolanti, non solo per le loro dimensioni, ma anche per l’elevata risoluzione di misurazione e l’alta densità della nuvola di punti finale. La maggior parte delle scene sono ottenute attraverso l’integrazione di diverse nuvole di punti, acquisite con diversi sensori (fotocamere e laser scanner) e piattaforme (UAV, *Unmanned Aerial Vehicle*).

I laser scanner terrestri impiegati includono un FARO Focus 3D X 130 e 120 e un Riegl VZ-400. I rilievi fotogrammetrici del Sacro Monte di Varallo sono stati eseguiti con una Nikon D880E mentre per Bologna e Trento sono state impiegate rispettivamente una Nikon D3100 e D3X. Un UAV era equipaggiata con una SONY Ilce 5100L mentre il DJI Phantom 4 Pro ha la sua fotocamera integrata.

3.2 Pre-elaborazione dati

Le nuvole di punti raccolte sono state inizialmente pre-elaborate per rendere le strutture delle nuvole più omogenee (Tabella 1). La pre-elaborazione delle nuvole è stata eseguita in *CloudCompare* e ha seguito 3 passaggi:

- traslazione spaziale;
- sottocampionamento;
- scelta delle *features*.

La *traslazione spaziale* delle nuvole di punti è necessaria a causa della georeferenziazione delle scene: i valori delle coordinate sono infatti troppo grandi per essere elaborati dalla rete neurale, quindi le coordinate vengono troncate e ogni singola scena viene spostata spazialmente vicino all’origine del sistema (0,0,0).

L’operazione di *sottocampionamento* si è resa necessaria a causa dell’elevato numero di punti (per lo più ridondanti) in ogni scena (>20 milioni di punti). L’opzione del sottocampionamento casuale è stata scartata perché avrebbe limitato la ripetibilità del test, quindi sono stati testati altri due metodi: il sottocampionamento basato sullo spazio e sull’*octree*.

Dal confronto dei risultati derivanti dall’applicazione del sottocampionamento basato su *octree* e spazio, si è optato per la seconda impostazione. La variazione nei risultati del test è stata dell’1%, pertanto sono state preferite l’uniformità e la semplicità

ADDESTRAMENTO

Nome	Numero di punti	Scena	Acquisizione dati	Numero di Classi (esclusa “Altro”)	Sottocampionamento (cm)
1_TR_cloister	15,740,229	Interno/Esterno	TLS + UAV	8/9	1
2_TR_church	20,862,139	Interno	TLS	8/9	1
3_VAL_room	4,188,066	Interno	TLS	6/9	1
4_CA_church	4,850,807	Esterno	TLS + UAV	6/9	1
5_SMV_chapel_1	3,783,412	Esterno	TLS + UAV	9/9	1
6_SMV_chapel_2to4	6,326,871	Interno/Esterno	TLS + UAV	9/9	1
7_SMV_chapel_24	3,571,064	Esterno	TLS + UAV	9/9	1
8_SMV_chapel_28	3,156,753	Esterno	TLS + UAV	9/9	1
9_SMV_chapel_10	2,193,189	Interno/Esterno	TLS + UAV	6/9	1
10_SStefano_portico_1	3,783,699	Esterno	Fotogrammetria terrestre	8/9	1
11_SStefano_portico_2	10,047,392	Esterno	Fotogrammetria terrestre	8/9	1
12_KAS_pavillion_1	598,384	Interno/Esterno	TLS	4/9	1
13_KAS_pavillion_2	325,822	Interno/Esterno	TLS	4/9	1
14_TRE_square	10,045,227	Esterno	Fotogrammetria terrestre	8/9	1.5
15_OTT_church	13,264,040	Interno/Esterno	TLS	9/9	1.5

TEST

Nome	Numero di punti	Scena	Acquisizione dati	Numero di Classi (esclusa “Altro”)	Sottocampionamento (cm)
A_SMG_portico	16,165,924	Esterno	TLS + UAV	9/9	1
B_SMV_chapel_27to35	16,200,442	Interno/Esterno	TLS + UAV	9/9	1

Tabella 1. Caratteristiche principali dell’ArCH dataset. Il numero di classi in ogni scena potrebbe essere utile per scegliere la scena di validazione durante le fasi di addestramento.

date dall'utilizzo del metodo basato sulla scelta dello spazio minimo tra i punti. Per quanto riguarda lo spazio è stato impostato uno spazio minimo tra i punti di 0,01 m, in questo modo è garantito un elevato livello di dettaglio, ma allo stesso tempo è possibile ridurre notevolmente il numero di punti e la dimensione del *file*, oltre a regolarizzare la struttura geometrica della nuvola di punti.

Nel *framework* del DL, la selezione delle *features* è soggetta a due diversi approcci. Il primo consiste nel selezionare il minor numero possibile di *features* e lasciare che la rete neurale impari da esse. Il secondo, utilizzato principalmente per dataset più piccoli, prevede la selezione di specifiche *features* "artigianali", estratte *ad hoc*, in modo da facilitare l'apprendimento delle reti e migliorare le prestazioni complessive, seppur aumentando i tempi computazionali. In questo caso, la maggior parte delle *features* sono solitamente estratte per compiti specifici (Zhang et al. 2019) e possono essere suddivise e classificate in intrinseche ed estrinseche, o anche utilizzate per descrittori locali e globali (Han et al., 2018; Weinmann et al. al., 2015). Le caratteristiche locali definiscono le proprietà statistiche delle informazioni geometriche del vicinato locale, mentre le caratteristiche globali descrivono l'intera geometria della nuvola di punti. Le proprietà più utilizzate sono quelle locali, come i descrittori basati su autovalori (*eigenvalue*), *3D shape context*, ecc. Ciononostante, sono qui fornite solo *features* intrinseche comuni, al fine di consentire agli utenti di trovare le combinazioni più appropriate. Le uniche *features* calcolate sono le normali.

Le normali dei punti sono state calcolate su CloudCompare, il più delle volte con un modello di superficie locale piana e orientate con uno *spanning tree* minimo con $Knn=10$. L'orientamento delle normali è stato quindi verificato in MATLAB®.

La struttura delle nuvole di punti è x, y, z, r, g, b , classe di annotazione, N_x, N_y, N_z .

4. DEFINIZIONE DELLE CLASSI

Attraverso il riconoscimento automatico degli elementi architettonici, gli autori desiderano supportare e velocizzare il processo di ricostruzione delle geometrie 3D per i modelli HBIM. In questo contesto, è essenziale scegliere classi per il benchmark che siano già disponibili nel software *object-oriented* o negli standard su cui si basano. In questo modo le classi predette dalla rete neurale corrispondono esattamente alle categorie BIM e, una volta ricostruita la geometria, sarà possibile associare le sue informazioni direttamente alle classi specifiche.

Nello stato dell'arte, alcuni lavori hanno già associato la semantica, basata su tassonomie e ontologie, a elementi del patrimonio (Mallik e Chaudhury, 2012) o modelli HBIM (Quattrini et al., 2017; Yang et al., 2019). Tuttavia, non ci sono ancora studi che uniscano la semantica del dominio BIM con il riconoscimento automatico degli elementi architettonici attraverso le tecniche DL.

Oltre allo standard su cui si basa il dominio BIM (l'IFC), sono stati presi in considerazione altri standard, per avere una visione complessiva e multilivello dei beni culturali. In particolare, gli standard indagati sono quelli volti a descrivere:

- edifici e contesto circostante, come BOT (Building Topology Ontology) e CityGML (Geography Markup Language);
- i beni culturali stessi, come CIDOC-CRM (International Committee for Documentation- Conceptual Reference Model) e il dizionario AAT (Art and Architecture Thesaurus) del Getty Institute.

Organizzando semanticamente i dati, essi possono essere gestiti con un vocabolario comune e la suddivisione in classi non è quindi arbitraria, ma oggettiva e standardizzata, uguale per tutti gli utenti e riferita a un lessico già codificato. È stato così sviluppato un metodo unificato per la classificazione degli elementi architettonici (Malinverni et al., 2019).

Il concetto di Livello di Dettaglio (LOD) deriva dal modello dati CityGML e permette di descrivere un oggetto secondo differenti scale di rappresentazione, in cui sia le geometrie rappresentate che le informazioni inserite vanno dal generale al particolare. Questo concetto è stato quindi applicato alla segmentazione semantica delle nuvole di punti: dapprima si è cercato di capire a quale livello di dettaglio sono segmentate le nuvole di punti nello stato dell'arte e, successivamente, sono state individuate le classi corrispondenti negli standard citati.

In CityGML, il LOD 0 descrive una scala regionale e paesaggistica, il LOD 1 la regione o la città, il LOD 2 i quartieri della città, il LOD 3 e 4 i modelli architettonici rispettivamente con gli elementi esterni e interni.

Se consideriamo alcuni esempi in letteratura sulla classificazione delle nuvole di punti nel campo geospaziale utilizzando le reti neurali (Landrieu e Simonovsky, 2018; Hackel et al., 2017), possiamo affermare che il livello di dettaglio raggiunto fino ad ora è compreso tra LOD 1 e 2. Tra le classi standard maggiormente identificate (es. vegetazione, strade, edifici, ecc.), mancano ancora i singoli elementi architettonici.

È stata quindi definita l'annotazione semantica delle nuvole di punti secondo un LOD 3/4 di CityGML.

In particolare, nel CityGML, il LOD 3 prevede la realizzazione di un modello architettonico dettagliato e nel suo schema sono inseriti oggetti come porte e finestre.

Le classi identificate sono, all'interno di "Feature" *Boundary Surface*, "Pavimento", "Tetto" e "Muro" e, all'interno di "Feature" *Openings*, "Finestra" e "Porta".

Per quanto riguarda lo standard IFC, la categoria che contiene gli elementi architettonici è *IfcBuildingElement*, una sottoclasse di *IfcElement*. In questa categoria, diversi elementi architettonici possono descrivere un edificio, ma solo alcuni di questi sono comuni al dominio dei BCC, dal momento che altri sono troppo specifici per le nuove costruzioni o per alcune tecniche costruttive. Le classi individuate sono, quindi: "Colonna", "Porta", "Tetto", "Scala", "Muro" e "Finestra", due delle quali già in comune con il modello dati CityGML.

Inoltre, poiché le classi incluse in questi due standard non sono sufficienti per descrivere correttamente un bene architettonico, è stato preso in esame anche il vocabolario AAT al cui interno sono state selezionate le classi "Volte" e "Archi", da *Architectural e Structural elements*, e "Modanature" da *Surface elements*.

A seguito di alcuni studi e risultati di segmentazione con le *features* 3D (Grilli et al., 2019b), si è deciso di modificare la classificazione proposta in (Malinverni et al., 2019; Pierdicca et al., 2020), separando la classe delle "Colonne" da quella delle "Lesene" e inserendo queste ultime nella nuova classe "Modanature", dove sono presenti anche cornici e cornicioni.

A tal fine sono state selezionate 9 classi (Figura 3), più un'ulteriore definita "Altro", contenente tutti i punti non appartenenti alle classi precedenti (es. dipinti, altari, panchine, statue, grondaie...).

Queste classi sono state utilizzate per l'annotazione delle nuvole di punti (Figura 5). Tuttavia, è prevista la possibilità di estendere ulteriormente questo schema per un livello di dettaglio più elevato (LOD 4/5), da sfruttare per la segmentazione semantica delle istanze (*Instance segmentation*). I lettori interessati possono approfondire questo argomento in (Mo et al., 2019).

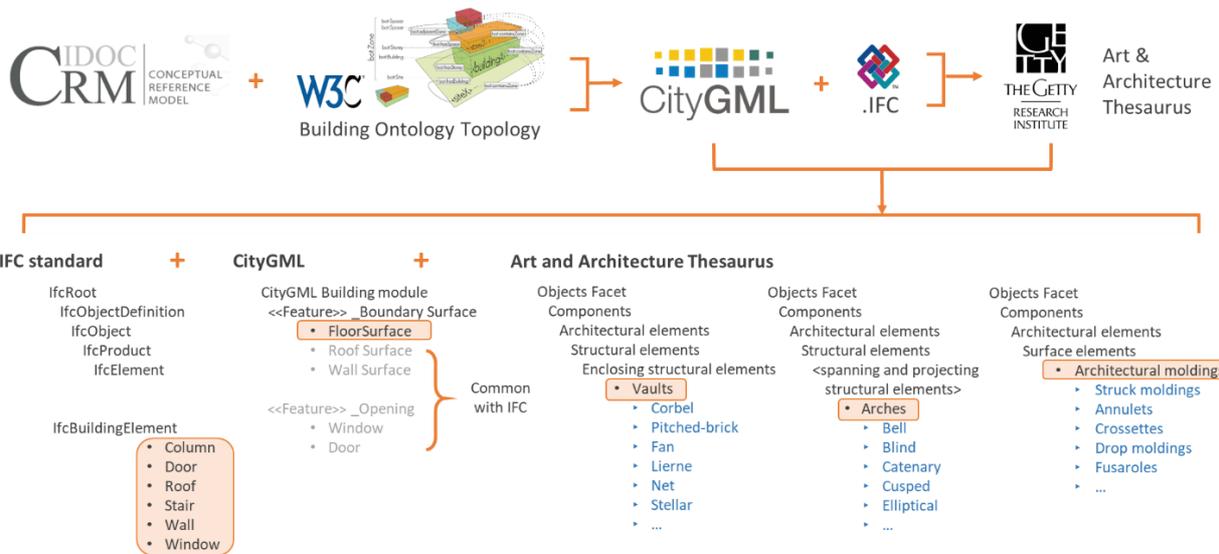


Figura 3. Classi selezionate per l'annotazione delle nuvole di punti.

5. OBIETTIVI DEL BENCHMARK E VALUTAZIONE

Il benchmark è suddiviso in due sezioni:

- la prima con le nuvole di punti già annotate per la fase di training;
- la seconda con le nuvole di punti per la fase di test e valutazione.

In questo modo, il benchmark proposto potrebbe essere utilizzato per addestrare e valutare lo stato dell'arte e nuovi metodi di classificazione e segmentazione semantica. Inoltre, gli utenti hanno la possibilità di scegliere arbitrariamente le scene utili ai loro scopi. L'attività di benchmark offre anche una valutazione delle *performance* dei metodi di segmentazione proposti dai ricercatori. Se gli autori invieranno i risultati predetti dalle reti, questi verranno confrontati automaticamente con quelli della *ground truth* e verranno pubblicati sul sito in termini di accuratezza complessiva, *F1 score*, precisione, *recall* e *Intersection over Union* (IoU).

$$\text{Accuracy} = \frac{\text{Number of correct predictions}}{\text{Total number of predictions}} \quad (1)$$

$$\text{Precision} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Positives (FP)}} \quad (2)$$

$$\text{Recall} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Negatives (FN)}} \quad (3)$$

$$\text{F1 score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

$$\text{IoU} = \frac{|A \cap B|}{|A \cup B|} = \frac{|I|}{|U|} \quad (5)$$

Attualmente, sono riportate online le *performance* di alcune reti dello stato dell'arte: PointNet (Qi et al., 2017a), PointNet ++ (Qi et al., 2017b), la Point Convolutional Neural Network (PCNN) (Atzmon et al., 2018), la Dynamic Graph Convolutional Neural

Network (DGCNN) (Wang et al., 2019) e la DGCNN-Modificata (Pierdicca et al., 2020). Queste reti sono state valutate usando in fase di addestramento 10 scene su 15 disponibili, tuttavia sono stati anche pubblicati i risultati derivanti dall'utilizzo della DGCNN con *features* 3D su tutto l'ArCH dataset (Matrone et al., 2020).

	N. scene per l'addestram.	Acc.	Prec	Rec.	F1Score
DGCNN	10	0.733	0.721	0.733	0.707
PointNet++	10	0.441	0.480	0.487	0.448
PointNet	10	0.307	0.405	0.306	0.287
PCNN	10	0.623	0.642	0.608	0.636
DGCNN-Mod	10	0.743	0.748	0.742	0.722
DGCNN-Mod	15	0.837	0.829	0.837	0.823
DGCNN-Mod+3Dfeat	15	0.865	0.853	0.864	0.856

Tabella 2. Metriche dello stato dell'arte con reti addestrate con parte dell'ArCH dataset e con l'intero dataset. La scena di test utilizzata in questo caso è *B_SMV_chapel_27to35*.

Una questione critica da menzionare è lo sbilanciamento delle classi (Figura 4). Alcune di esse, infatti, sia per le fasi di addestramento che test, hanno un numero di punti maggiore e questo può influenzare negativamente le prestazioni della rete e le varie metriche. Andranno pertanto adottati accorgimenti (come l'utilizzo della *focal loss*) per attenuare questa problematica.

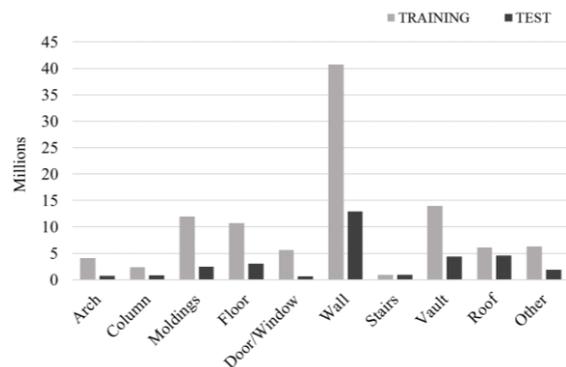


Figura 4. Numero di punti per classe.



Figura 5. Esempio di nuvole di punti annotate dall'ArCH dataset.

5.1 Linee guida per l'annotazione da parte degli utenti

Come precedentemente detto, nel dataset sono anche disponibili per gli utenti i punti che non sono ancora state annotate, ma che potrebbero essere annotate dagli utenti stessi e poi ricaricate sul sito per tutta la comunità scientifica.

Vista l'eterogeneità degli elementi architettonici, sono però state definite alcune linee guida per l'annotazione delle nuvole di punti. Queste linee guida consentono ad altri ricercatori di contribuire all'espansione del dataset. Quest'ultimo è stato manualmente annotato con un comune software di elaborazione delle nuvole di punti come *CloudCompare*. Tuttavia, nella pagina del benchmark, è disponibile per gli utenti anche uno strumento di annotazione basato sull'applicazione web di un editor di segmentazione semantica.

Considerando ogni classe, escluse quelle standard di muri, pavimenti, tetti e scale, le linee guida seguite per l'annotazione sono state:

- **Colonne.** In questa classe sono state inserite solo colonne o pilastri indipendenti, sia a sezione circolare che quadrata. Come accennato in precedenza, le lesene o le semicolonne addossate alle pareti sono state incluse nella classe "Modanature".
- **Modanature.** In questa classe sono stati inclusi gli stucchi e tutti gli altri tipi di modanature, come cornici delle finestre e delle porte o cornicioni, oltre alle lesene e alle semicolonne precedentemente citate (Figura 6). Più in generale, tutto ciò che sporge dalla muratura rientra in questa classe.



Figura 6. Esempio di modanature (viola) e colonne (rosso).

- **Porte e finestre** sono state combinate in un'unica classe, dato il numero ridotto di punti e la loro geometria simile.
- **Volte.** Ogni tipo di volta (a botte, crociera, cupola...) è stato incluso in questa classe. Se le singole volte sono interrotte da archi sporgenti rispetto alla volta stessa allora devono essere suddivise nelle classi "Arco" e "Volta", altrimenti tutti i punti rientrano nella classe "Volta".
- **Archi.** Questa classe comprende sia gli archi della facciata sia quelli che dividono una volta dall'altra, ma solo se aggettanti (Figura 7).
- **Altro.** Tutto ciò che non rientra nelle classi precedenti è stato qui incluso. Questa classe ha il solo scopo di raggruppare alcuni elementi architettonici o di arredo (pluviali, panchine, balaustre...) che potrebbero essere utili in futuro e che, allo stesso tempo, aiutano nella comprensione generale della nuvola di punti. Per le fasi di addestramento e test, si consiglia di escludere questa classe, in quanto potrebbe influenzare negativamente la

loss function, le prestazioni generali delle reti neurali o qualsiasi altro algoritmo utilizzato.



Figura 7. Interno chiesa con esempio di archi aggettanti (blu) rispetto all'altezza delle volte (arancione).

6. CONCLUSIONI

Questo contributo descrive il benchmark ArCH, ideato per la segmentazione semantica della nuvola di punti 3D. La piattaforma fornisce ai ricercatori milioni di punti, annotati secondo uno standard definito, insieme a un quadro di valutazione generalizzato. Il set di dati comprende nuvole di punti sia annotate che non annotate e gli autori invitano la comunità di ricerca a contribuire a questo compito difficile, ma essenziale.

Si spera che in futuro il benchmark diventi la fonte di riferimento per testare e condividere nuovi risultati e *framework* verso un'automazione del riconoscimento degli oggetti per architetture complesse come quelle dei BBCC.

Alcuni studi precedenti hanno dimostrato che le reti neurali offrono strategie affidabili per la classificazione dei dati 3D dei BBCC. È giusto affermare che, al contrario di altri ambiti di ricerca, i BBCC presentano ancora diversi colli di bottiglia e proprio a causa di questi, fino ad ora, non è ancora emerso un metodo che assicuri prestazioni di gran lunga migliori rispetto agli altri. Fornendo un dataset aperto e codici *open source*, gli autori si auspicano di poter definire una linea comune per le implementazioni future, nella misura in cui verranno sviluppati nuovi algoritmi in un prossimo futuro.

Lo sbilanciamento delle classi, l'eterogeneità degli elementi architettonici e la complessità delle scene sono attualmente le principali criticità e le questioni ancora aperte.

Siamo fiduciosi che il benchmark soddisfi le esigenze delle attività di ricerca nel campo dei beni architettonici e diventi una risorsa centrale per lo sviluppo di metodi nuovi, efficienti e accurati per la classificazione del patrimonio 3D. Il benchmark contribuirà fortemente ad arricchire il corpo di conoscenze per la segmentazione semantica dei BBCC attraverso metodi automatici e supervisionati basati sull'apprendimento.

RINGRAZIAMENTI

Gli autori desiderano ringraziare Ilaria Bonfanti, Valeria De Ruvo, Emanuele Pontoglio e Gloria Rizzo del Laboratorio di Geomatica del Politecnico di Torino per il loro aiuto nell'annotazione delle nuvole di punti.

Un sincero ringraziamento va, inoltre, al VR Lab (DISEG) e al G4CH Lab del Politecnico di Torino (DAD) e all'"Ente Gestore dei Sacri Monti" con il suo direttore dott. Elena De Filippis per aver consentito alla pubblicazione delle proprie nuvole di punti. Grazie infine alle curie per aver gentilmente accettato di condividere le nuvole di punti degli edifici religiosi.

BIBLIOGRAFIA

- Armeni, I., Sener O., Zamir, A. R., Jiang, H., Brilakis, I., Fischer, M., Savarese, S., 2016. 3D semantic parsing of large-scale indoor spaces, *Proc. IEEE CVPR*, pp. 1534–1543.
- Atzmon, M., Maron, H., Lipman, Y., 2018. Point convolutional neural networks by extension operators. *arXiv:1803.10091*.
- Bello, S. A., Yu, S., Wang, C., 2020. Review: deep learning on 3D point clouds. *arXiv:2001.06280v1*.
- Bruno, N., Roncella, R. A., 2018. Restoration oriented HBIM system for Cultural Heritage documentation: The case study of Parma cathedral. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.*, 42, pp. 171-178.
- Bitelli, G., Dellapasqua, M., Girelli, V., Sanchini, E., Tini, M., 2017. 3D Geomatics Techniques for an integrated approach to Cultural Heritage knowledge: The case of San Michele in Acerboli's Church in Santarcangelo di Romagna. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.*, 42, pp. 291-296.
- Wang, C.; Dai, Y.; Elsheimy, N.; Wen, C.; Retscher, G.; Kang, Z.; Lingua, A. Progress on ISPRS benchmark on multisensory indoor mapping and positioning, *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.*, 42, pp. 1709-1713.
- CloudCompare, 2019. CloudCompare 3D point cloud and mesh processing software Open Source Project. cloudcompare.org (last access: 01/05/2020).
- Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S. and Schiele, B., 2016. The cityscapes dataset for semantic urban scene understanding. In *Proc. IEEE CVPR*, pp. 3213-3223.
- De Deuge, M., Quadros, A., Hung, C., Douillard, B., 2013. Unsupervised feature learning for classification of outdoor 3D scans. In: *Australasian Conference on Robotics and Automation*, Vol. 2, p. 1
- Geiger, A., Lenz, P., Stiller, C., Urtasun, R., 2013. Vision meets robotics: The KITTI dataset. *International Journal of Robotics Research*, 32(11), pp. 1231–1237.
- Griffiths, D., Boehm, J., 2019. A Review on Deep Learning Techniques for 3D Sensed Data Classification. *Remote Sensing*, 11(12):1499
- Grilli, E., Özdemir, E., & Remondino, F., 2019a. Application of Machine and Deep Learning strategies for the classification of heritage point clouds. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLII-4/W18, pp. 447-454.
- Grilli, E., Farella, E. M., Torresani, A., and Remondino, F., 2019b. Geometric features analysis for the classification of cultural heritage point clouds. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLII-2/W15, pp. 541–548.
- Hackel, T., Savinov, N., Ladicky, L., Wegner, J.D., Schindler, K. and Pollefeys, M., 2017. Semantic3D. net: A new large-scale point cloud classification benchmark. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. IV-1-W1, pp. 91-98.
- Han, X., Jin, J.S., Xie, J., Wang, M., Jiang, W., 2018. A comprehensive review of 3D point cloud descriptors. *ArXiv*, abs/1802.02297.
- Kharroubi, A., Hajji, R., Billen, R., Poux, F., 2019. Classification and integration of massive 3D points clouds in a Virtual Reality (VR) environment. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLII-2/W17, pp. 165-171.
- Korc, F. & Förstner, W., 2009. eTRIMS Image Database for interpreting images of man-made scenes. Dept. of Photogrammetry, University of Bonn, *Tech. Rep.* TR-IGG-P-2009-01.
- Landrieu, L., Simonovsky, M., 2018. Large-scale point cloud semantic segmentation with superpoint graphs. *Proceedings of Proc. IEEE CVPR*, pp. 4558-4567.
- Llamas, J., Leronés, P. M., Medina, R., Zalama, E., Gómez-García-Bermejo, J., 2017. Classification of Architectural Heritage Images Using Deep Learning Techniques. *Applied Sciences*, 9, Volume 7, p. 992.
- Malinverni, E.S., Pierdicca, R., Paolanti, M., Martini, M., Morbidoni, C., Matrone, F., Lingua, A., 2019. Deep learning for semantic segmentation of point clouds. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* Vol. XLII-2/W15, pp. 735–742.
- Mallik, A., Chaudhury, S., 2012. Acquisition of multimedia ontology: an application in preservation of cultural heritage. *International journal of multimedia information retrieval*, (1)4, 249-262.
- Matrone, F., Grilli, E., Martini, M., Paolanti, M., Pierdicca, R., Remondino, F., 2020. Comparing Machine and Deep Learning Methods for Large 3D Heritage Semantic Segmentation. *ISPRS Int. J. Geo-Inf.*, 9, 535.
- Mo, K., Zhu, S., Chang, A. X., Yi, L., Tripathi, S., Guibas, L. J., & Su, H., 2019. Partnet: A large-scale benchmark for fine-grained and hierarchical part-level 3D object understanding. *Proc. IEEE CVPR*, pp. 909-918.
- Munoz, D., Bagnell, J. A., Vandapel, N., Hebert, M., 2009. Contextual classification with functional max-margin markov networks. *Proc. IEEE CVPR*, pp. 975–982.
- Murtiyoso, A.; Grussenmeyer, P., 2019. Point Cloud Segmentation and Semantic Annotation Aided by GIS Data for Heritage Complexes. In *Proceedings of the 8th International Workshop 3D-ARCH “3D Virtual Reconstruction and Visualization of Complex Architecture”*, Bergamo, Italy, 6–8 February 2019, pp. 523–528.
- Murtiyoso, A., Grussenmeyer, P., 2020. Virtual Disassembling of Historical Edifices: Experiments and Assessments of an Automatic Approach for Classifying Multi-Scalar Point Clouds into Architectural Elements. *Sensors*, 20(8), 2161.
- Özdemir, E., Toschi, I., Remondino, F., 2019. A multi-purpose benchmark for photogrammetric urban 3D reconstruction in a controlled environment. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, Vol. XLII-1/W2, pp. 53–60.
- Paolanti, M., Romeo, L., Martini, M., Mancini, A., Frontoni, E., Zingaretti, P., 2019. Robotic retail surveying by deep learning visual and textual data. *Robotics and Autonomous Systems*, 118, pp. 179-188.
- Pierdicca, R., Paolanti, M., Matrone, F., Martini, M., Morbidoni, C., Malinverni, E.S., Frontoni, E., Lingua, A.M., 2020. Point Cloud Semantic Segmentation Using a Deep Learning Framework for Cultural Heritage. *Remote Sens.*, 12, 1005.
- Quattrini, R., Pierdicca, R., Morbidoni, C., 2017. Knowledge-based data enrichment for HBIM: Exploring high-quality models using the semantic-web. *Journal of Cultural Heritage*, 28, pp. 129-139.

Qi, C.R., Su, H., Mo, K., Guibas, L.J., 2017a. Pointnet: Deep learning on point sets for 3D classification and segmentation. In Proc. IEEE CVPR, pp. 652–660.

Qi, C.R., Yi, L., Su, H., Guibas, L.J., 2017b. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *arXiv:1706.02413*.

Roynard, X., Deschaud, J.E. and Goulette, F., 2018. Paris-Lille-3D: A large and high-quality ground-truth urban point cloud dataset for automatic segmentation and classification. *The International Journal of Robotics Research*, 37(6), pp.545-557.

Semantic Segmentation Editor: web labeling tool for camera and LIDAR data. Available online: <https://github.com/Hitachi-Automotive-AndIndustry-Lab/semantic-segmentation-editor>. (last access on 25/04/2020).

Serna, A., Marcotegui, B., Goulette, F. and Deschaud, J.-E., 2014. Paris-rue-madame database: a 3d mobile laser scanner dataset for benchmarking urban detection, segmentation and classification methods. *Proc. 4th ICPRAM Conference*.

Teboul, O., Kokkinos, I., Simon, L., Koutsourakis, P., Paragios, N., 2013. Parsing Facades with Shape Grammars and Reinforcement Learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 7, Volume 35, pp. 1744-1756.

Tyleček, R. and Šára, R., 2013, September. Spatial pattern templates for recognition of objects with regular structure. In *German Conference on Pattern Recognition*, pp. 364-374. Springer, Berlin, Heidelberg.

Vallet, B., Brédif, M., Serna, A., Marcotegui, B. and Paparoditis, N., 2015. TerraMobilita/iQmulus urban point cloud analysis benchmark. *Computers & Graphics*, 49, pp.126-133.

Wang, Y., Sun, Y., Liu, Z., Sarma, S.E., Bronstein, M.M., Solomon, J.M., 2019. Dynamic graph CNN for learning on point clouds. *ACM Trans. Graph. arXiv:1801.07829v2*.

Wang, C., Hou, S., Wen, C., Gong, Z., Li, Q., Sun, X., Li, J., 2018. Semantic Line Framework-based Indoor Building Modeling using Backpack Laser Scanning Point Cloud. *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 143, pp. 150-166.

Weinmann, M., Jutzi, B., Hinz, S., & Mallet, C., 2015. Semantic point cloud interpretation based on optimal neighborhoods, relevant features and efficient classifiers. *ISPRS Journal of Photogrammetry and Remote Sensing*, 105, pp. 286-304.

Wu, Z., Song, S., Khosla, A., Yu, F., Zhang, L., Tang, X., Xiao, J., 2015. 3Dshapenets: A deep representation for volumetric shapes. *Proc. IEEE CVPR*, pp. 1912– 1920.

Yang, X., Lu, Y.C, Murtiyoso, A., Koehl, M., Grussenmeyer, P., 2019. HBIM modeling from the surface mesh and its extended capability of knowledge representation. *ISPRS International Journal of Geo-Information*, (8)7, 301.

Zhang, K., Hao, M., Wang, J., Silva, C.W., & Fu, C., 2019. Linked Dynamic Graph CNN: Learning on Point Cloud via Linking Hierarchical Features. *ArXiv*, abs/1904.10014.

INTELLIGENZA ARTIFICIALE E REALTÀ AUMENTATA PER LA CONDIVISIONE DEL PATRIMONIO CULTURALE

R. Spallone^a, V. Palma^b

^a Dipartimento di Architettura e Design, Politecnico di Torino, viale Mattioli 39, 10125 Torino, Italy – roberta.spallone@polito.it

^b Dipartimento di Architettura e Design, Politecnico di Torino, viale Mattioli 39, 10125 Torino, Italy – Torino, Italy -
valerio.palma@polito.it

PAROLE CHIAVE: Intelligenza artificiale, realtà aumentata, *free and open source software*, patrimonio culturale, patrimonio architettonico, condivisione, salvaguardia, valorizzazione

KEY WORDS: Artificial intelligence, augmented reality, free and open source software, cultural heritage, architectural heritage, sharing, safeguard, enhancement

RIASSUNTO

Le problematiche relative alla fruizione e condivisione del Patrimonio Culturale sempre più coinvolgono l'integrazione tra tecnologie digitali per l'analisi, l'interpretazione, la comunicazione, la diffusione. L'intreccio fra *artificial intelligence* (AI) e *augmented reality* (AR) si sta rivelando particolarmente interessante nella valorizzazione del patrimonio, grazie alle nuove possibilità di riconoscimento e ancoraggio fra prodotti digitali e consistenze fisiche, in altre parole tra virtuale e reale. L'idea che guida le differenti esperienze di seguito presentate è che libri, documenti, manufatti, edifici, città, nella loro consistenza fisica, possano essere una chiave d'accesso condivisa alla conoscenza del bene, per orientarsi tra piani di analisi e livelli informativi che si moltiplicano. Le scelte rispetto agli strumenti digitali impiegati sono state orientate a criteri di sostenibilità dei progetti e dei loro sviluppi, preferendo *free and open source software* (FOSS), o comunque *software* gratuito, per contenere i costi di implementazione e favorire l'interoperabilità dei risultati. Gli esperimenti che seguono, di carattere spiccatamente euristico, illustrano come, proprio a partire dallo spazio fisico, siano stati realizzati strumenti di espansione delle esperienze finalizzate alla condivisione del patrimonio e, in prospettiva, alla partecipazione della collettività alla sua salvaguardia.

ABSTRACT

Problems related to the use and sharing of Cultural Heritage increasingly involve the integration of digital technologies for analysis, interpretation, communication and dissemination. Artificial intelligence (AI), augmented reality (AR) and their links are proving particularly interesting in the enhancement of heritage, thanks to new possibilities of recognition and anchoring between digital products and physical features, that is between virtual and real. The experiences presented below are based on the idea that books, documents, artefacts, buildings, cities, as physical assets, can be a shared access key to knowledge on heritage, in order to navigate through the spread of information layers and levels of analysis. The choice of the employed digital tools was based on sustainability criteria for project development and scaling, favoring free and open source software (FOSS) or otherwise free to use software, to reduce implementation costs and improve output interoperability. In the described experiments, which have a marked heuristic character, starting right from the physical space, tools have been developed to enhance the experience of sharing cultural heritage and subsequently foster collective participation and safeguard actions.

1. INTRODUZIONE

Le problematiche relative alla fruizione e condivisione del Patrimonio Culturale sempre più coinvolgono l'integrazione tra tecnologie digitali per l'analisi, l'interpretazione, la comunicazione, la diffusione.

Nello specifico, *artificial intelligence* (AI) e *augmented reality* (AR) si stanno rivelando particolarmente interessanti nella valorizzazione del patrimonio, grazie alle nuove possibilità di riconoscimento e ancoraggio fra prodotti digitali e consistenze fisiche, in altre parole tra virtuale e reale.

Nel campo dei Beni Culturali, il "diluvio di dati" (Anderson, 2008) riguarda in particolare le acquisizioni digitali, i modelli interpretativi e le elaborazioni spaziali, la modellazione informativa, nonché la digitalizzazione di archivi esistenti. Le sfide sulla valorizzazione di questa conoscenza riguardano la sua lettura, la diffusione e la permanenza del tempo. L'idea perseguita è che lo spazio fisico del costruito possa essere una chiave d'accesso condivisa alla conoscenza del bene, per

orientarsi tra piani di analisi e livelli informativi che si moltiplicano.

Nel seguito dell'articolo verrà descritta una serie di esperienze di carattere spiccatamente euristico, basate su esperimenti di introduzione, nei Beni Culturali e in sistemi già esistenti di gestione dell'informazione digitale ad essi relativa, di tecnologie digitali per il riconoscimento di oggetti, strumenti che, pur cominciando a relazionarsi con lo spazio costruito, hanno più spesso trovato affermazione in altri ambiti. Le esperienze descritte sono nate dal confronto e dagli intrecci fra gli specifici interessi di conoscenza dei due autori e dagli sviluppi di ricerche con altri studiosi.

2. SALVAGUARDIA E CONDIVISIONE DEL CULTURAL HERITAGE

Come si è sopra ricordato, l'entità, la quantità e l'eterogeneità di dati di conoscenza, di analisi, di interpretazione, raccolti,

elaborati e immediatamente disponibili vengono prodotti dalla collettività per le più disparate finalità. A questo si affianca l'altrettanto ampia quantità di beni materiali e immateriali che costituiscono un patrimonio culturale diffuso ed esteso, dal passato al presente (dunque analogici e digitali) il cui valore in molti casi è rappresentato dalla sommatoria di singoli beni che divengono sistema, anziché dall'eccezionalità di un singolo bene. Proprio per queste caratteristiche, tale patrimonio risulta spesso a rischio di obsolescenza. Il problema del riconoscimento di valore e della conservazione di tali beni si intreccia con le necessità della memoria, della comunicazione e della diffusione.

Non solo, la produzione di dati da parte degli utenti può entrare nel sistema virtuoso di un'informazione circolare. Un archivio multimediale e facilmente accessibile può infatti essere la base per un sistema di raccolta di informazioni *crowdsourced*, ovvero collezionate dagli utenti stessi del servizio, che comprendano documenti (ad esempio fotografie), dati di fruizione (per guidare manutenzione, organizzazione di servizi), *link* tra le diverse voci (per affinità di contenuti e caratteristiche, o in base agli interessi comuni degli utenti). Simili modalità di interazione con gli strumenti permettono di accrescere il patrimonio conoscitivo della collettività, suggerendo inedite interpretazioni, visioni, connessioni.

I modelli digitali possono quindi assumere il ruolo di efficaci database per raccogliere e sintetizzare sia le analisi degli studiosi, sia i *feedback* degli utenti. Inoltre, è stato osservato che tali modelli offrono possibilità di visualizzazione altrimenti impossibili nella realtà tangibile, che permettono l'integrazione dei frammenti, l'ispezione degli oggetti in tutte le loro coordinate fisiche: insomma, accrescono gli strumenti di analisi, di ricerca, di studio, ma allo stesso tempo anche quelli per la comunicazione delle informazioni (Irace, 2013).

3. AI E AR: TECNOLOGIE, METODOLOGIE E NESSI

Sistemi come audio-guide, pannelli informativi o codici QR già permettono di accedere a informazioni rilevanti all'interno di siti culturali. Tuttavia, queste soluzioni richiedono progetti *ad hoc* e possono prevedere infrastrutture costose, deperibili, poco inclini all'aggiornamento dei contenuti e a modalità di consultazione che si adattino a utenti con esigenze diverse.

Alcune tecnologie permettono però a strumenti diffusi come gli *smartphone* di connettere una posizione nello spazio, un'opera d'arte, un oggetto architettonico e le sue caratteristiche morfologiche a una selezione specifica di contenuti digitali.

L'AR mette a disposizione metodi intuitivi ed efficaci per l'accesso alle informazioni che hanno una relazione con lo spazio, abilitando l'interazione con oggetti virtuali in maniera più simile all'interazione col mondo reale.

I sistemi di *machine learning*, applicati ad esempio al riconoscimento di immagine, possono connettersi a estese basi di dati e reperire informazioni su un oggetto inquadrato con il dispositivo, o sulla categoria di oggetti a cui appartiene.

Una delle sfide che simili opportunità permettono di affrontare è l'ottimizzazione delle risorse investite nei beni culturali, poiché forniscono soluzioni a basso costo, facilmente scalabili, che valorizzano sia l'esperienza diretta di opere e siti, sia la ricerca e la documentazione che li raccontano.

In particolare, le *mobile computing technologies* si prestano a soluzioni sia per grandi siti molto frequentati, in cui fornire servizi adeguati al volume e alla varietà dell'utenza può richiedere interventi complessi e costosi, sia per siti minori e al

di fuori di *network* turistici consolidati, nei quali sorveglianza e manutenzione mancano spesso di fondi adeguati.

Negli esperimenti descritti, le applicazioni di AR e AI permettono di superare alcuni dei limiti delle soluzioni precedentemente adottate per fini analoghi, valorizzando l'esperienza dei siti culturali e rendendo le collezioni digitali più accessibili.

L'AR è una tecnologia che può rendere 'esplicite' delle informazioni associate a un oggetto fisico, attraverso la sovrapposizione alle immagini del mondo reale, ottenute tramite una videocamera, ad altre informazioni elaborate digitalmente, sull'interfaccia di un unico dispositivo e in tempo reale. Come sintetizzato da Amin e Govilkar (2015) i sistemi di AR si compongono delle tre fasi di *recognition*, *tracking* e *mixing* (Fig. 1). Diversi approcci basati su tecniche visuali permettono di individuare oggetti anche senza l'applicazione di specifici *marker* di riconoscimento, e di tracciare le posizioni relative dell'oggetto osservato e dell'osservatore (Yang et al., 2013; Younes et al., 2017). Questi strumenti sono oggi alla portata di molti dispositivi mobili, anche dotati di singola telecamera, grazie a diversi *software development kit* (SDK) per lo sviluppo di *app*. che permettono il riconoscimento e il tracciamento non solo di immagini, ma anche di oggetti tridimensionali, consentendo il diffondersi applicazioni che dialogano con uno spazio sempre più ampio e complesso.

Anche l'intelligenza artificiale (AI) – e più nello specifico il *deep learning* (DL) applicato alla *computer vision*, ovvero l'interpretazione automatica di immagini – è un campo tecnologico che interagisce sempre più da vicino con uno spazio di scala architettonica e urbana. Questo legame si è rafforzato in particolare con l'interesse nello sviluppo di auto a guida autonoma (Cordts et al., 2016). Tuttavia, in questi casi si tratta soprattutto di immagini del piano stradale, mentre sono ancora pochi gli studi che approfondiscono l'interpretazione dello spazio costruito, dell'architettura, della forma della città storica e contemporanea (Stathopoulou e Remondino, 2019).

Il DL (Goodfellow et al., 2016) è una classe di algoritmi di *machine learning* il cui recente sviluppo è legato alla potenza di calcolo oggi disponibile, nonostante gli strumenti siano basati su modelli elaborati sin dagli anni '40 del secolo scorso. Una spinta particolare allo sviluppo di questi algoritmi per la *computer vision* è arrivata con la disponibilità di grandi *data set* di immagini, e con lo sviluppo dei motori grafici, requisiti indispensabili per rendere efficienti i processi di *addestramento* dei modelli.

Quando il patrimonio culturale è incarnato dai beni architettonici, emerge la necessità che questi siano documentati attraverso acquisizione e trattamento rigorosi dei dati metrici, alla base dell'informazione spaziale delle applicazioni di tecnologie AR e AI. Un riferimento fondativo in questo senso risale al lavoro di Portales et al. (2009) che presenta un'applicazione AR a basso costo, basata su un software open-source, creata per mostrare la volta barocca sopra l'altare maggiore della Cattedrale di Valencia, di recente smantellata. L'esperimento si fonda su una documentazione metrica che integra la misurazione con laser scanner terrestre (TLS) ad alta risoluzione con la registrazione delle informazioni sulle *texture* e i dati radiometrici. Ricerche più recenti connettono i prodotti delle acquisizioni mediante TLS e *structure from motion* (SfM) con la comunicazione del patrimonio attraverso AR. È il caso dei lavori sul teatro Romano di Fano (Quattrini et al., 2016), in cui la tecnologia di AR viene utilizzata per visualizzare il modello virtuale ottenuto attraverso TLS, sul Santuario di Loreto, la cui digitalizzazione è collegata attraverso l'AR con il Libro I di Bramante (Clini et al., 2017) e sulla chiesa di

Sant'Antonio abate (RC), con l'acquisizione mediante TLS e aeromobile a pilotaggio remoto (UAV) del manufatto finalizzata alla visualizzazione *in situ* degli interni, mediante AR (Barrile et al., 2018).

Gli esperimenti che seguono illustrano una serie di applicazioni a casi studio di differente natura, dal materiale documentario e archivistico, agli apparati effimeri, al patrimonio architettonico e urbano e intendono mostrare la linea evolutiva della ricerca.

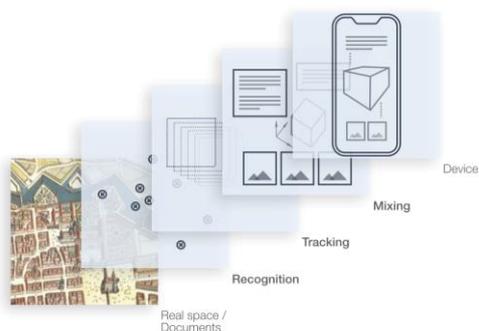


Figura 1. Schema delle fasi di recognition, tracking e mixing nelle applicazioni di AR basate su tracciamento di immagine.

4. ESPERIMENTI

4.1 Iconografie

Il primo esperimento, proposto nell'ambito del Convegno Internazionale FORTMED, tenutosi a Torino nel 2018, ha inteso applicare le potenzialità della AR per accrescere il dato informativo della mappa di Torino presente nel *Theatrum Sabaudiae* (1682). Precedenti studi sulle relazioni tra le mura, la Cittadella e le porte, avevano condotto alla raccolta di numerose iconografie e prodotto diversi disegni digitali interpretativi e modelli 3D, relativi all'andamento delle mura nelle differenti espansioni della città, al tracciamento e dimensioni della Cittadella e alle porte della città di impianto romano e barocco, trasformate o demolite (Spallone, 2017a; Spallone, 2017b; Vitali, 2015; Vitali, 2017).

L'associazione di documenti e fonti tradizionali con differenti elaborazioni in formato digitale ha richiesto nuove modalità di comunicazione rispetto agli archivi fisici e il supporto di media appropriati. Inoltre, ha posto il problema della gestione e dell'accessibilità alla quantità e qualità di dati oggi a disposizione.

A questo fine è stato sviluppato un prototipo di applicazione di AR per dispositivi mobili che ha permesso di interagire con la mappa del *Theatrum Sabaudie*, al fine di condividere con gli studiosi e il pubblico le fonti più importanti e i risultati delle precedenti indagini. I principali obiettivi perseguiti sono stati:

- rendere la mappa del *Theatrum* un primo punto di accesso alla ricerca svolta sulle tavole del libro;
- offrire a chi consulta il documento originale o riprodotto in formato cartaceo uno strumento di conoscenza approfondita;
- rendere espliciti i collegamenti tra i contenuti iconografici del *Theatrum*, e tra gli stessi e altre fonti;
- studiare l'interoperabilità e la scalabilità di soluzioni che uniscano archivio fisico e virtuale;
- valorizzare gli archivi fisici, attraverso la possibilità di avvicinare gli utenti ai documenti offerta da un sistema di servizi informativi in continuo sviluppo.

Gli strumenti digitali impiegati in questa esperienza sono stati scelti in base a criteri di sostenibilità del progetto e dei suoi

sviluppi, preferendo *free and open source software* (FOSS), o comunque *software* gratuito, per contenere i costi di implementazione e favorire l'interoperabilità dei risultati (Scianna et al., 2016). Il progetto di AR è stato sviluppato con l'impiego del *software* per la creazione di videogiochi Unity, integrato all'SDK Vuforia. I due programmi sono prodotti commerciali, ma offrono gratuitamente funzionalità complete per le fasi di sviluppo (esclusa distribuzione). Le applicazioni prodotte con Unity sono installabili su diverse piattaforme per dispositivi mobili. Nell'applicazione sviluppata, Vuforia gestisce i meccanismi di *recognition* e di *tracking* di immagini predefinite, ovvero la ricostruzione in uno spazio tridimensionale della posizione dell'oggetto visualizzato. Le immagini sono registrate in un *database* contenuto nell'applicazione e sono individuate da un algoritmo attraverso *features* facilmente distinguibili. Queste immagini *target*, per essere efficacemente riconosciute e tracciate, devono rispondere a specifiche caratteristiche, quali un buon contrasto e l'assenza elementi geometrici ripetitivi che possono essere tra loro confusi. Una volta stimata la posizione dell'immagine *target*, Unity gestisce la fase di *mixing*. Il sistema di AR colloca un modello tridimensionale nello spazio virtuale, secondo una posizione definita in fase di programmazione in relazione a quella del *target*, e aggiornata dinamicamente. Anche i modelli tridimensionali sono memorizzati nell'applicazione.

Il riconoscimento dell'immagine, realizzato sulla riproduzione a colori, si è dimostrato efficace sia sull'immagine a colori che su quella al tratto, permettendo il funzionamento dell'*app* con le due versioni (Fig. 2). La mappa della città contenuta nel *Theatrum Sabaudiae* è infatti pubblicata da un originale del volume a colori, di cui una copia è custodita presso l'Archivio Storico della Città di Torino, e in un volume da incisioni su rame.

L'applicazione è stata completata con un *menu* da cui consultare una lista di immagini corrispondenti a documenti correlati al monumento osservato. Con questa funzione si è sperimentata una possibile estensione della rete di *link* tra documenti d'archivio e informazione digitale. Lo strumento si presenta quindi come un accesso interattivo e immediato a una lista di riferimenti. Le potenzialità di questa rete informativa sono state approfondite integrando all'applicazione la comunicazione con l'archivio digitale Cult (Bortot et al., 2017; Cecchini et al., 2018), un *database* geo-spaziale dotato di interfacce *online* per la gestione e di un servizio *web* per l'accesso ai dati.



Figura 2. Visualizzazione in realtà aumentata del modello tridimensionale della Porta Palatina. L'immagine target è una riduzione in scala della mappa del *Theatrum Sabaudiae*. Immagine tratta da Palma et al., 2018.

4.2 Scenografie

Il secondo esperimento è consistito nell'elaborazione di applicazioni di AR per l'espansione dell'esperienza teatrale. Il progetto ha unito studiosi di Ispanistica e Tecniche Digitali di Rappresentazione con la Fondazione TPE (Teatro Piemonte Europa) e l'associazione Saveria Project. Quest'ultima, ha operato la riscrittura teatrale del Don Quixote di Cervantes per uno spettacolo tenuto dal 28 al 30 marzo 2019 nella Biblioteca Storica "Arturo Graf" dell'Università di Torino (Fig. 3), nell'ambito del programma Creative Europe (Culture): Q. THEATRE-Theatrical recreations of Don Quixote in Europe. Il concetto di Rappresentazione, come "spettacolo" e come "immagine" ha connesso il lavoro delle figure coinvolte, in un progetto che ha integrato la produzione teatrale con inedite esperienze di AR. Si è trattato dunque di reinterpretare il movente del romanzo, ossia il confronto tra un uomo e una tecnologia rivoluzionaria: il libro a stampa, come confronto tra uomo e nuovi media digitali.

Lo spettatore ha avuto accesso in AR a riproduzioni audio/video in streaming, a immagini digitali e modelli 3D che hanno esteso l'esperienza teatrale fino all'interazione con video, immagini e scenografie immersive. A tal fine è stata prodotta un'app per dispositivi mobili (*tablet* e *smartphone*) chiamata Dulcinea.

La sfida condotta dal gruppo multidisciplinare ha ambito a innescare nello spettatore riflessioni più profonde sull'essenza stessa di questa realtà "aumentata", così come Cervantes ha inteso portare il lettore a ragionare sul significato della strana e peculiare follia di chi crede incondizionatamente in quello che legge.

L'applicazione Dulcinea ha tracciato un percorso nella biblioteca le cui tappe sono state continuamente interrotte dagli interventi *live* di quattro attori chiamati a "fare qualcosa" sul Chisciotte e presentati dalla voce in cuffia come "le espressioni corporee della mia incorporeità".

L'AR ha consentito di visualizzare in tempo reale figure e modelli virtuali in corrispondenza di immagini di oggetti presenti nella biblioteca (scaffali, libri, poster, schermi...). Come nell'esperimento precedente, sono state sfruttate principalmente le funzioni di *recognition* e *tracking* di immagini.

Oltre ad abilitare l'accesso all'AR, l'applicazione ha assunto la funzione di riprodurre una voce narrante che ciascuno spettatore ha potuto ascoltare tramite auricolari. Gli spettatori sono stati accompagnati nelle diverse stanze dagli attori e guidati dalla voce in cuffia. Si è creata per il pubblico una doppia interazione: con lo spazio scenico e con i dispositivi digitali il



Figura 3. Particolare della "scenografia" in AR realizzata nell'ambito del programma Creative Europe (Culture): Q. THEATRE-Theatrical recreations of Don Quixote in Europe. Il target è una porzione di libreria, assimilata a immagine bidimensionale. Immagine tratta da Scamuzzi et al., 2019.

cui uso all'interno dello spettacolo ha simulato quello quotidiano, di acquisizione, invio, fruizione e condivisione di informazioni e prodotti multimediali svolto in maniera individuale anche in mezzo a molte altre persone. Così, la capacità propria dello strumento di costruire mondi virtuali e singolari ha permesso di enfatizzare il contrasto con lo spazio reale e con gli interventi degli attori.

Nelle diverse stanze lo spettatore è stato invitato a esplorare l'ambiente attraverso la videocamera, attivando diversi effetti di AR quali: l'alterazione animata di immagini bidimensionali, l'apparizione di modelli tridimensionali e l'avvio di video.

Anche in questo caso, l'app è stata programmata usando il motore Unity e installata su dispositivi con sistemi Android e iOS, mentre lo strumento impiegato per le funzioni di AR è Vuforia. Il plug-in AudioStream per Unity è stato usato per ricevere via streaming l'audio della voce narrante.

Per il broadcasting dell'audio sono stati adottati software open source di streaming, quali Icecast 2 e *broadcast using this tool*.

4.3 Architettura

Una terza esperienza ha approfondito la possibilità di interagire con le informazioni spaziali digitali in un ambiente di scala architettonica, sfruttando la diffusione di tecnologie di AR per dispositivi mobili che permettono il tracciamento di oggetti basati su modelli 3D piuttosto che immagini.

Il lavoro si è innestato su una ricerca dedicata agli atri barocchi torinesi, descritta in Spallone e Vitali (2017). Queste opere, caratterizzate da sistemi voltati complessi, sono un caso unico nell'architettura del tempo. Gli atri sono realizzati seguendo un particolare schema spaziale che presenta spazi unitari con assenza di sostegni intermedi e volte composite in muratura.

Sono tre i principali tipi di sistema voltato classificati: le volte "stellari", le volte "Planteriane" (dal nome del loro inventore Gian Giacomo Plantery) e le volte "a fascioni". Sono stati individuati oltre settanta atri rispondenti a queste caratteristiche, progettati da architetti noti a livello internazionale – come Guarini, Juvarrà, Vittone – e meno note figure che hanno contribuito alla definizione e alla diffusione di queste tipologie – come Garove, Baroncelli e Plantery.

Il valore, riconosciuto da critici e studiosi, di questo bene culturale contrasta con la scarsa attenzione rivolta da un pubblico più vasto, in particolare nella considerazione dell'insieme degli atri come un sistema urbano.

La ricerca ha prodotto una catalogazione di questi spazi che ha riferito ad una mappa le opere entro le mura della Torino di impianto romano e barocco, ha determinato le dimensioni degli atri e le proprietà geometriche dei sistemi voltati, e ha riportato il corpo di riferimenti bibliografici e archivistici connessi. Alcuni esempi paradigmatici sono stati inoltre approfonditi attraverso disegni bidimensionali e modelli interpretativi 3D (Fig. 4).

L'idea perseguita è stata rendere disponibili i risultati della ricerca sugli atri barocchi torinesi all'interno degli stessi spazi studiati, sfruttando la flessibilità crescente delle tecnologie di AR e ritenendo l'insieme degli atri un valido campo di prova, data l'omogeneità di alcune caratteristiche spaziali – si tratta di ambienti concavi, di dimensioni contenute, con superfici spesso decorate e simili condizioni di illuminazione.

Abbiamo condotto una rassegna degli strumenti di AR disponibili sul mercato e scelto per un primo test quelli del kit di sviluppo ARKIT di Apple. Quando questo esercizio è stato avviato, nei primi mesi del 2019, diversi sistemi commerciali iniziavano a offrire funzioni per il riconoscimento e il tracciamento di oggetti tridimensionali. Tuttavia, la maggior

parte degli strumenti era indirizzata ad applicazioni tabletop, ossia per l'ancoraggio degli strati digitali ad oggetti di piccole dimensioni. Il sistema scelto, come altre soluzioni di model tracking, ha richiesto la costruzione di nuvole di punti rade, come modelli di riferimento per l'ancoraggio. In diversi SDK questa operazione può essere realizzata attraverso app di supporto per la scansione di oggetti. Nel nostro caso, abbiamo impiegato una demo prodotta da Apple, modificata per consentire il rilevamento a una maggiore distanza.

L'uso dell'app di scansione è una procedura guidata che porta ad inquadrare l'oggetto di riferimento da diversi punti di vista, identificando i *feature points*, ovvero punti omologhi riconoscibili attraverso la sequenza di immagini della telecamera – scelti tra i soli punti contenuti in una *bounding box*, a sua volta in AR. Al termine del rilevamento, il *software* permette di verificare che l'oggetto rilevato sia riconosciuto e tracciato, e di posizionare manualmente l'origine del sistema di riferimento dello spazio virtuale costruito.

La nuvola di punti può quindi essere esportata nell'app prodotta per la visualizzazione in AR dei materiali informatici, nella quale sono inseriti modelli da visualizzare in corrispondenza degli oggetti riconosciuti. In particolare, abbiamo testato la sovrapposizione di quattro modelli geometrici prodotti nella ricerca sugli atri torinesi, corrispondenti alle volte dei palazzi Carignano (Guarino Guarini, 1679), Novarina (Gian Giacomo Plantery, primo decennio del XVII sec.), Cigliano (Plantery, 1707-08) e Coardi di Carpenetto (Amedeo di Castellamonte, non datato).

Le scansioni si sono rivelate più efficaci quando limitate a porzioni limitate di spazio e ricche di dettagli plastici, come capitelli e imposte delle volte (Fig. 5). La disponibilità, nell'SDK, di funzioni di "tracciamento esteso", basate su sensori di movimento e analisi visiva della scena, ha permesso comunque di ottenere la visualizzazione stabile *in situ* dell'intero modello geometrico di volta – anche quando la porzione di target risulta fuori dall'inquadratura – con un'accuratezza della posizione compatibile con le finalità e con il dettaglio dei modelli di ricostruzione realizzati. Il progetto non ha implementato un'interazione con dati accessibili via *web*, ma ha studiato la possibilità di inserire le informazioni sugli atri nella piattaforma Cult, già usata in progetti precedenti, stabilendo un punto di partenza per la scalabilità del progetto in futuri sviluppi. Tali sviluppi includono la strutturazione di percorsi tematici attraverso il centro storico di Torino come mezzo di scoperta e condivisione delle opere e dei materiali della ricerca, per cittadini e turisti

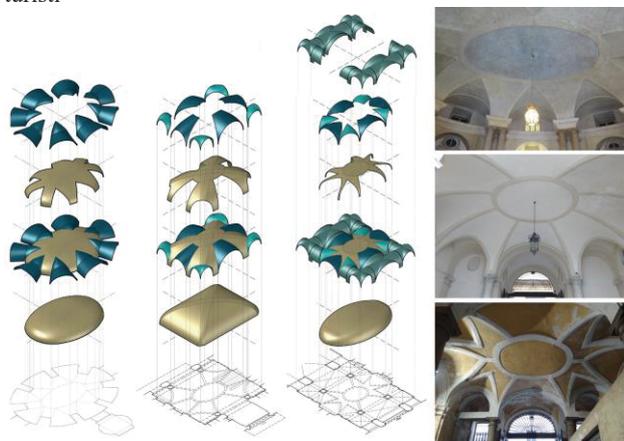


Figura 4. Sistemi voltati complessi negli atri barocchi torinesi: Palazzo Carignano, Palazzo Cigliano, Palazzo Novarina. Immagine tratta da Palma et al., 2019b.

L'interazione dell'AR con altri siti e con una più ampia dimensione urbana include sfide tecniche e concettuali sull'organizzazione dell'informazione spaziale e sull'accessibilità, non solo rispetto ai dati, ai metadati e ai documenti da consultare, ma anche per quanto riguarda la gestione di grandi *set* di modelli di riconoscimento e ancoraggio per rendere l'informazione fruibile secondo le modalità "aumentate" che stiamo studiando.

4.4 Città

Il quarto progetto, ancora in corso, sviluppa l'idea di costruire un sistema per accedere in maniera intuitiva e rapida a un *database* di documenti multimediali, con testi, immagini, modelli tridimensionali, video e informazioni strutturate. Si tratta di archivi spesso già resi disponibili e in espansione nell'ambito dell'*heritage* – come visto nei progetti di ricerca precedenti e con la piattaforma Cult. Il progetto è portato avanti dal Centro Interdipartimentale FULL del Politecnico di Torino. Nei primi esiti della ricerca, è stata realizzata un'app che mette in comunicazione spazio fisico e spazio digitale usando tecnologie di DL per la *computer vision*.

I primi esperimenti hanno riguardato un prototipo per l'Area Archeologica Centrale di Roma, che include i Fori Imperiali, dove è forte l'esigenza pratica di un servizio informativo preciso e completo che possa essere fornito in lingue, a livelli di dettaglio e per tipi di visitatore diversi, ma con impatti limitati sul sito e sulla manutenzione. Il sistema sviluppato si compone di due parti principali. La prima è l'archivio digitale, accessibile via *web*, geo-spaziale, multimediale, per il quale è stato impiegata la piattaforma Cult, predisposta attraverso una campagna di raccolta ed elaborazione di contenuti informativi. La seconda parte è l'AI. Il modello impiegato appartiene alla classe delle reti neurali convoluzionali (*convolutional neural networks*, CNN), largamente impiegate in molti campi per gli specifici compiti della *computer vision* (Hosny et al., 2018; Webb, 2018). Il modello esegue il riconoscimento di immagine ed è installato direttamente sul dispositivo per funzionare anche senza una connessione a internet – necessaria solo per il primo *download* delle informazioni dal *database* e per l'eventuale aggiornamento. L'applicazione, attraverso un'interfaccia semplice ed intuitiva, permette di inquadrare i monumenti usando la telecamera e, a riconoscimento avvenuto, presenta un'etichetta con il nome relativo. Cliccando sulla schermata del riconoscimento, è avviata una *query* sul *database* locale; quindi, l'app mostra una seconda schermata con le informazioni sul monumento, come testi, immagini, dati schematici, il *link* a una mappa con la localizzazione e modelli 3D navigabili (Fig. 6).

Il prototipo principale funziona su piattaforme iOS, ed è stato sviluppato in ambiente Xcode usando il linguaggio SWIFT e la libreria ML Core per il *machine learning*. Un più semplice prototipo per Android è stato sviluppato in partenariato con l'università di Roma Tre, per testare la compatibilità con il modello di CNN. Quest'ultimo è sviluppato in Python attraverso le librerie Keras e TensorFlow per il DL. Come base è stato usato un modello MobileNet, particolarmente adatto allo sviluppo per dispositivi mobili (Howard et al., 2017).

All'attuale stato di sviluppo, l'app è in grado di riconoscere 46 monumenti che presentano una varietà di stati di conservazione, dimensioni, caratteri spaziali e punti di vista per gli osservatori. Per il *training* dell'AI è stato raccolto un *data set* basato su 50-100 fotografie per ogni monumento, appositamente scattate.

Le immagini sono state selezionate per ottenere una rappresentazione esaustiva del monumento, ponendo specifica



Figura 5. Applicazioni di AR basate su *model tracking*. Dalla colonna sinistra: Palazzo Novarina, Palazzo Coardi di Carpenetto, Palazzo Cigliano. Le immagini superiori mostrano le fasi di produzione della nuvola di punti per l'ancoraggio. Immagine tratta da Palma et al., 2019b.

attenzione alle viste più comuni per i visitatori e includendo dettagli, viste panoramiche e diverse condizioni di illuminazione (diurne).

Il *data set* è stato quindi ampliato a 600 immagini per monumento, attraverso tecniche di *augmentation* basate sulla manipolazione di immagine, e in particolare con l'uso della libreria Augmentor per il linguaggio Python. L'*augmentation* limita l'insorgenza di *overfitting* (l'aderenza eccessiva del modello di DL al *set* di dati di input, che ne impedirebbe il corretto funzionamento) nelle situazioni in cui la quantità di dati a disposizione per il *training* è ridotta (Perez e Wang, 2017). Il processo applica alle immagini originali trasformazioni come *cropping*, rotazioni, ribaltamenti, aggiunta di rumore e alterazioni del colore, ed è di documentata efficacia per i modelli di riconoscimento di immagine (Goodfellow et al.,

2016; Perez e Wang, 2017; Bloice et al., 2019). Delle 600 immagini ottenute, 100 sono state inserite in un *data set* per la validazione dei risultati.

Il *training* è stato infine condotto usando un cluster di GPU Nvidia Tesla. Il *training* per le 46 classi selezionate, effettuato attraverso 1000 *epochs* (le iterazioni attraverso cui l'allenamento della rete neurale ne altera i parametri sulla base del *data set* di *training*) ha permesso di ottenere una accuratezza del 99,88%, misurata sul *data set* di validazione (percentuale di immagini classificate correttamente).

Il modello risultante è stato verificato sul campo, dove si è dimostrato affidabile, pur con un'accuratezza minore di quella misurata sul *data set* di validazione, indicando che gli avanzamenti del lavoro dovranno apportare ulteriore varietà alla *data set* di input (per esempio raccogliendo immagini in diverse

condizioni meteorologiche). Inoltre, il comportamento eterogeneo del sistema di riconoscimento rispetto ai diversi edifici ha suggerito, nelle fasi di correzione del *data set* e del modello, la suddivisione degli “oggetti” più grandi e complessi (e.g. la *Domus Augustana* sul Palatino) in più *classi* di riconoscimento che indirizzano alla stessa voce nel *database* di informazioni.

La rete neurale costruita occupa 13MB di spazio su disco, il che la rende ideale per il funzionamento, anche *offline*, sui dispositivi mobili.

La flessibilità degli strumenti e anche la struttura essenziale del sistema ne consentono la rapida estensione a nuovi siti, ed è già stata sviluppata un’applicazione per la città di Torino chiamata *ARCH-I Torino*, in grado di riconoscere circa 80 monumenti ed opere d’arte, già distribuita gratuitamente sul mercato digitale. In questo caso l’app sfrutta il *database* di MuseoTorino.it, un’infrastruttura tesa a realizzare un “museo diffuso della città” che raccoglie migliaia di schede informative dettagliate su monumenti e opere d’arte, prima accessibili solo attraverso un sito *web* e strumenti di *query* testuale o mappe.

Queste sperimentazioni nel campo dell’AI si sono basate sull’operazione ancora concettualmente semplice di associare alle immagini i nomi propri dei monumenti. L’interazione delle tecnologie di DL e delle discipline che si occupano dello spazio costruito può però prendere in analisi, in luogo dei singoli oggetti, il riconoscimento di categorie, ad esempio tipologie edilizie, caratteri costruttivi e stilistici, qualità dell’abitare (climatiche, visive, relative ai flussi) e altre forme di segmentazione dei materiali studiati. Alcuni di questi approfondimenti sono già in corso (cfr. Stathopoulou e Remondino, 2019), ma si può sottolineare la necessità ulteriori studi, nonché di *data set* per il *training*, che stimolino lo specifico sviluppo di modelli di DL per l’architettura, i beni culturali e l’interpretazione dello spazio urbano.

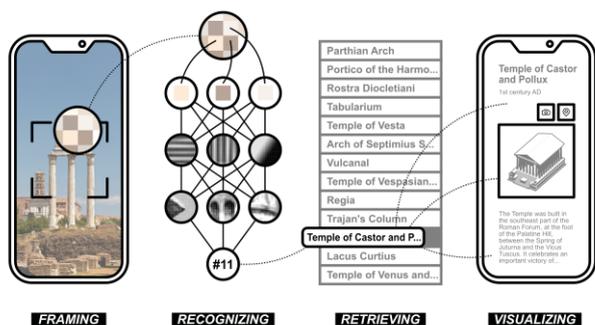


Figura 6. Schema del funzionamento dell’app per il riconoscimento di monumenti tramite DL. Immagine rielaborata da Andrianaivo et al. 2019 (fotografia del tempio di Castore e Polluce: hl_1001 CC BY-NC 2.0).

5. CONCLUSIONI

Nelle diverse esperienze descritte emerge la comune sfida dell’interazione tra tecnologie e conoscenza del patrimonio culturale e architettonico. Una questione trasversale riguarda il modo in cui organizziamo, classifichiamo, visualizziamo le informazioni sullo spazio urbano: quali astrazioni utilizziamo, perché, e come queste sono definite. Attraverso tecnologie come quelle dell’AI, sempre più dispositivi sono in grado elaborare quantità enormi di informazioni sullo spazio costruito e sul patrimonio culturale. Resta, però, a chi svolge ricerca in questi campi il compito di definire le categorie, le classi, le “ontologie” che costituiscono le griglie concettuali entro cui ordinare i

materiali e con cui indirizzare la nostra conoscenza. Il tema della chiarezza e dello scopo delle astrazioni, delle convenzioni, degli standard è tanto più rilevante quanto è maggiore il volume di dati che siamo in grado di produrre, dati che spesso portano a confondere la quantità di informazioni con la conoscenza. Le ICT offrono possibilità di fruizione inedite, tra cui sistemi di visualizzazione interattivi e soluzioni per ricerche rapide e complesse. Non solo: i nuovi strumenti permettono di definire nuove reti di relazioni tra i documenti, e di ridefinirle dinamicamente in base ad esigenze e contenuti in aggiornamento. Gli esperimenti condotti studiano questa topologia dell’informazione estesa dagli archivi digitali, cercando e sottoponendo a *test* il ruolo dello spazio fisico come punto di riferimento per il coordinamento di informazioni e strumenti. L’AI, l’AR, i loro nessi e il rapporto con l’informazione spaziale si presentano come occasioni per riflettere sull’interpretazione dello spazio come chiave per la gestione e per l’accesso all’informazione digitale.

6. RINGRAZIAMENTI

Le sezioni 3, 4.3 e 4.4 sintetizzano ed elaborano gli esiti delle ricerche pubblicate in Palma et al., 2018, Palma, 2019, Palma et al., 2019a, Palma et al. 2019b, Andrianaivo et al., 2019.

La sezione 4.1 sintetizza ed elabora gli esiti della ricerca pubblicata in Palma et al. (2018). Il progetto Cult, citato nella sezione, è stato sviluppato presso il dipartimento ICEA dell’Università di Padova e coordinato da Luigi Stendardo e Andrea Giordano.

Il paragrafo 4.2 sintetizza ed elabora gli esiti della ricerca pubblicata in Scamuzzi et al. (2019).

Il progetto descritto nella sezione 4.4 è stato condotto presso il Centro FULL | The Future *Urban Legacy* Lab del Politecnico di Torino, da un team composto da Matteo Robiglio, Claudio Casetti, Francesca Frassoldati, Louis N. Andrianaivo e Valerio Palma, in partnership con Roberto d’Autilia. Désirée Adiutori, del Dipartimento di Scienze Computazionali dell’Università di Roma Tre ha contribuito allo sviluppo del prototipo dell’app per piattaforme Android.

Il presente articolo di cui gli autori hanno condiviso l’organizzazione, gli obiettivi e l’impostazione metodologica, è stato redatto da Roberta Spallone (sezioni 1, 2, 4.1, 4.2), Valerio Palma (sezioni 3, 4.3, 4.4, 5).

7. BIBLIOGRAFIA

- Amin, D., Govilkar, S. (2015). Comparative Study of Augmented Reality SDK’s. *International Journal on Computational Science & Applications*, 5(1), 11–26.
- Anderson, C. (2008). The End of Theory: The Data Deluge Makes the Scientific Method Obsolete. *Wired*.
- Andrianaivo, L. N., D’Autilia, R., Palma, V. (2019). Architecture recognition by means of convolutional neural networks. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2-W15, 77–84.
- Barrile, V., Fotia, A., Bilotta, G. (2018). Geomatics and augmented reality experiments for the cultural heritage. *Applied Geomatics*, 10, 569–578.
- Bloice, M. D., Roth, P. M., Holzinger, A. (2019). Biomedical image augmentation using Augmentor. *Bioinformatics*.

- Bortot, A., Boscaro, C., Cecchini, C., Francesca, C., Cundari, M. R., Palma, V., Panarotto, F., Siviero, L. (2017). TU-CULT. Rivelazioni architettoniche nelle chiese di Santa Giustina e di Santa Maria dei Servi a Padova. In: Territories and frontiers of representation. 911–920.
- Cecchini, C., Cundari, M. R., Palma, V., Panarotto, F. (2019). Data, Models and Visualization: Connected Tools to Enhance the Fruition of the Architectural Heritage in the City of Padova. In C. L. Marcos (Ed.), *Graphic Imprints* (pp. 633–646). Springer International Publishing.
- Clini, P., Frontoni, E., Quattrini, R., Perdicca, R. (2017) Real/Not Real: Pseudo-Holography and Augmented Reality Applications for Cultural Heritage. In Ippolito, A., Cigola, M. (Eds.), *Handbook of Research on Emerging Technologies for Digital Preservation and Information Modeling* (pp. 201-227). IGI Global.
- Goodfellow, I., Bengio, Y., Courville, A. (2016). *Deep Learning*. MIT Press.
- Hosny, A., Parmar, C., Quackenbush, J., Schwartz, L. H., & Aerts, H. J. W. L. (2018). Artificial intelligence in radiology. *Nature Reviews Cancer*, 18(8), 500–510.
- Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., Adam, H. (2017). MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. *ArXiv:1704.04861 [Cs]*.
- Irace, F. (2013). The Animated Archive. In: Irace, F., Leyla Ciagà, G. (eds.) *Design & Cultural Heritage. Archivio Animato / Animated Archive*. Milano, Electa, pp. 10-14.
- Palma, V. (2019). Towards deep learning for architecture: A monument recognition mobile app. *ISPRS - International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2/W9, 551–556.
- Palma, V., Lo Turco, M., Spallone, R., Vitali, M. (2018). Augmented Iconography. AR applications to the fortified Turin in the *Theatrum Sabaudiae. Defensive Architecture of the Mediterranean*, IX, 1053–1060.
- Palma, V., Spallone, R., Vitali, M. (2019a). Digital Interactive Baroque Atria in Turin: A Project Aimed to Sharing and Enhancing Cultural Heritage. In Y. Tang, Q. Zu, & J. G. Rodríguez García (Eds.), *Human Centered Computing* (pp. 314–325).
- Palma, V., Spallone, R., Vitali, M. (2019b). Augmented Turin baroque atria: AR experiences for enhancing cultural heritage. *ISPRS - International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2/W9, 557–564.
- Perez, L., Wang, J. (2017). The effectiveness of data augmentation in image classification using deep learning. *arXiv preprint arXiv:1712.04621*.
- Portalés, C., Lerma, J. L., Pérez, C. (2009). Photogrammetry and augmented reality for cultural heritage applications. *The Photogrammetric Record*, 24(128), 316–331.
- Quattrini, R., Pierdicca, R., Frontoni, E., Barcaglioni, R. (2016). Virtual reconstruction of lost architectures: from the TLS survey to AR visualization. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume XLI-B5, 383-390.
- Scamuzzi, I., Moretti, S., Spallone, R., Lo Turco, M., Palma, V. (2019). Augmented Don Quixote: Performance teatrali e nuovi media digitali / Augmented Don Quixote: Theatrical performances and new digital media. *RIFLESSIONI. L'arte Del Disegno. Il Disegno Dell'arte / I REFLECTIONS. The Art of Drawing. The Drawing of Art*, 1, 1367–1374.
- Scianna, A., Gristina, S., Sciortino, R. (2016). Integrazione di sistemi GIS FOSS e modelli dati 3D PDF per la fruizione multimediale di beni monumentali e archeologici: Il Castello di Maredolce a Palermo. In F. Stanco & G. Gallo (Eds.), *Proceedings of ArqueoFOSS: Free, libre and open source software e open format nei processi di ricerca archeologica*. Archaeopress.
- Spallone, R. (2017a) The 'Regular Fortress' by Guarini and the Citadel of Turin. *Nexus Network Journal*, 19 (2), 255-277.
- Spallone, R. (2017b) City Gates. Proportional criteria and shape models for the design of Baroque gates in Turin. In: Echarri Iribarren, V. (ed.) *Defensive Architecture of the Mediterranean XV to XVIII Centuries. Vol. 5: Proceedings of FORTMED – Modern Age Fortification of the Mediterranean Coast*, 26-28 October, Alicante. Alicante, Universitat d'Alacant, pp. 143-150.
- Theatrum Sabaudiae. Teatro degli Stati del Duca di Savoia [1682]* (2000). Torino, Archivio Storico della Città di Torino.
- Spallone, R. and Vitali, M. (2017). Star-shaped and Planterian Vaults in Turin Baroque Atria. *Araene*.
- Stathopoulou, E.-K., Remondino, F. (2019). Semantic photogrammetry – Boosting image-based 3d reconstruction with semantic labeling. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2/W9, 685–690.
- Vitali, M. (2015) Digital 3D reconstruction for the multiscale investigation on the Drawing of the fortified wall of Turin. In: Rodríguez-Navarro, P. (ed.) *Defensive Architecture of the Mediterranean XV to XVIII Centuries. Vol. 1: Proceedings of FORTMED – Modern Age Fortification of the Western Mediterranean Coast*, 15-17 October 2015, València. València, Editorial Universitat Politècnica de València, pp. 329-336.
- Vitali, M. (2017) Digital models for the virtual reconstruction and the representation of the existing: the city gates of Turin. In: Echarri Iribarren, V. (ed.) *Defensive Architecture of the Mediterranean XV to XVIII Centuries. Vol. 5: Proceedings of FORTMED – Modern Age Fortification of the Mediterranean Coast*, 26-28 October, Alicante. Alicante, Universitat d'Alacant, pp. 489-496.
- Webb, S. (2018). Deep learning for biology. *Nature*, 554(7693), 555–557.
- Yang, M.-D., Chao, C.-F., Huang, K.-S., Lu, L.-Y., Chen, Y.-P. (2013). Image-based 3D scene reconstruction and exploration in augmented reality. *Automation in Construction*, 33, 48–60.
- Younes, G., Asmar, D., Elhadj, I., Al-Harithy, H. (2017). Pose tracking for augmented reality applications in outdoor archaeological sites. *Journal of Electronic Imaging*, 26(1).

APPROCCI DEEP LEARNING SU NUVOLE DI PUNTI SINTETICHE PER LA SEGMENTAZIONE SEMANTICA DI EDIFICI MONUMENTALI STORICI

LEARNING FROM SYNTHETIC POINT CLOUD DATA FOR HISTORICAL BUILDINGS SEMANTIC SEGMENTATION

C. Morbidoni^a, R. Pierdicca^b, M. Paolanti^a, R. Quattrini^b, R. Mammoli^b

^a Dipartimento di Ingegneria dell'Informazione, Università Politecnica delle Marche, 60100 Ancona, Italy; c.morbidoni@univpm.it (C.M.); m.paolanti@univpm.it (M.P.)

^b Dipartimento di Ingegneria Civile, Edile e dell'Architettura, Università Politecnica delle Marche, 60100 Ancona, Italy; r.pierdicca@univpm.it (R.P.); r.quattrini@univpm.it (R.Q.); r.mammoli@univpm.it (R.M.)

PAROLE CHIAVE: Segmentazione semantica, nuvole di punti, dataset sintetici, apprendimento profondo

KEY WORDS: Semantic segmentation, points clouds, synthetic dataset, deep learning

RIASSUNTO

Nello scenario del patrimonio storico, la segmentazione semantica delle nuvole di punti 3D sta ottenendo sempre più attenzione, poiché permette di riconoscere automaticamente gli elementi architettonici storici. Tuttavia, la segmentazione semantica è particolarmente interessante nell'architettura storica e classica, a causa della complessità delle forme e della limitata ripetibilità di elementi in edifici diversi, che rende difficile definire modelli comuni all'interno della stessa classe architettonica. Inoltre, poiché i modelli di apprendimento profondo richiedono una quantità considerevolmente grande di dati annotati per essere addestrati al fine di gestire adeguatamente nuove scene, la mancanza di (grandi) nuvole di punti annotate pubblicamente disponibili nel dominio degli edifici storici è un problema enorme, tanto da costituire un problema aperto nella ricerca. D'altra parte, la creazione di una massa critica di nuvole di punti etichettate mediante annotazione manuale è molto dispendiosa in termini di tempo e poco pratica. A tal fine, in questo lavoro viene esplorata l'idea di sfruttare i dati sintetici delle nuvole di punti per addestrare modelli di apprendimento profondo che eseguono la segmentazione semantica delle nuvole di punti. Per raggiungere questo scopo, viene presentata una versione migliorata della Dynamic Graph CNN (DGCNN) denominata RadDGCNN. Negli esperimenti presentati in questo lavoro, i modelli vengono addestrati su dataset sintetici (disponibili pubblicamente) di due diversi edifici storici: il Palazzo Ducale a Urbino e il Palazzo Ferretti ad Ancona, entrambi in Italia, acquisiti tramite Laser Scanner Terrestre. La rete RadDGCNN produce buoni risultati, dimostrando migliori prestazioni di segmentazione sui dataset reali.

ABSTRACT

In the historical heritage scenario, semantic segmentation of 3D point clouds is gaining more and more attention, since it might help to automatically recognize historical architectural elements. However, semantic segmentation is particularly challenging in historical and classical architecture, due to the shapes complexity and the limited repeatability of elements across different buildings, which makes it difficult to define common patterns within the same class. Furthermore, as deep learning models requires a considerably large amount of annotated data to be trained and tuned in order to properly handle unseen scenes, the lack of (big) publicly available annotated point clouds in the historical building domain is a huge problem, blocking the research in this direction. On the other hand, creating a critical mass of annotated point clouds by manual annotation is very time-consuming and impractical.

For this purpose, in this work we explore the idea of leveraging synthetic point cloud data to train deep learning models to perform semantic segmentation of point clouds obtained via Terrestrial Laser Scanning (TLS). To achieve this purpose, we present an improved version of the Dynamic Graph CNN (DGCNN) named RadDGCNN. In our experiments, we evaluate the trained models on synthetic dataset (publicly available) about two different historical buildings: the Ducal Palace in Urbino, Italy, and Palazzo Ferretti in Ancona, Italy. RadDGCNN yields good results, demonstrating improved segmentation performances on the TLS real datasets.

1. INTRODUZIONE

Nel contesto della gestione e conservazione del patrimonio culturale (CH), la creazione di rappresentazioni digitali accurate e ricche di edifici storici è di primaria importanza. Diversi lavori di ricerca studiano come ottenere informazioni accurate e affidabili su un edificio storico attraverso l'acquisizione della realtà e la modellazione 3D per supportare la documentazione dell'architettura ed il progetto di restauro. Il patrimonio storico richiede elaborazioni robuste per ottenere modelli HBIM (Heritage Building Information Modeling) che siano completamente interoperabili e ricchi del loro contenuto

informativo. La disponibilità di flussi di lavoro efficienti potrebbe rappresentare un passo molto importante verso una gestione più efficiente del patrimonio architettonico.

Il processo di documentazione del patrimonio parte solitamente dall'acquisizione di nuvole di punti 3D attraverso tecniche di scansione laser terrestre (TLS) o fotogrammetriche che producono nuvole di punti accurate e precise, fino a milioni di punti. I metodi di acquisizione fotogrammetrica e TLS consentono di generare una grande quantità di coordinate geometriche e 3D di una scena. Tuttavia, mentre le nuvole di punti sono utili per visualizzare una data scena, sono principalmente rappresentazioni non strutturate e carenti di

informazioni sui singoli oggetti, come superfici, dimensioni e semantica. Per questo motivo, le nuvole di punti sono spesso utilizzate, all'interno di ambienti software BIM dedicati, come punto di partenza per costruire rappresentazioni 3D parametriche, che incorporano informazioni semantiche e dove gli elementi architettonici vengono identificati e adeguatamente etichettati. Tale processo è spesso indicato come *Scan-to-BIM*.

La creazione di modelli 3D strutturati da nuvole di punti è un processo manuale che richiede tempo. Richiede agli esperti di dominio di ispezionare visivamente la nuvola di punti identificando diversi tipi di elementi architettonici di interesse, per poi modellare gli oggetti, posizionare correttamente ogni elemento sulla scena e annotarlo secondo i dizionari di dominio e/o le ontologie. Pertanto, vi è una forte necessità di automatizzare parzialmente il processo *Scan-to-BIM*, sviluppando metodologie e strumenti che siano in grado di assistere gli esperti del settore, rendendo l'intero processo più agile e veloce.

Un compito importante in questa direzione è quello di identificare automaticamente i diversi tipi di elementi architettonici nella nuvola di punti. Questa attività viene definita segmentazione semantica e implica la classificazione di ogni punto nella nuvola come appartenente a un particolare tipo di oggetto (ad esempio muro, tetto, colonna, volta, ecc.). Accelerare il processo *Scan-to-BIM*, attualmente eseguito manualmente, è cruciale, ma la segmentazione automatica dei dati 3D è ancora un agli albori per il Digital Cultural Heritage.

Mentre le tecniche di machine e deep learning si sono dimostrate efficaci nel segmentare semanticamente oggetti 2D (cioè immagini), nel caso di nuvole di punti 3D, essendo queste configurazioni geometriche di natura irregolare, caratterizzate dalla mancanza di una griglia ordinata, con un'elevata variabilità di densità, non strutturata e invariante alla trasformazione e alla permutazione, il compito è più difficile.

Recentemente, sono stati raggiunti risultati promettenti mediante l'applicazione di tecniche di deep learning specificamente progettate per gestire le nuvole di punti (Qi et al., 2017a; Qi et al. 2017b; Wang et al., 2019b).

Tali approcci sono stati applicati alla segmentazione di scene di interni (Armeni et al., 2016), cioè uffici, e valutati su benchmark standard, come Stanford 2D-3D-Semantics Dataset¹. Tuttavia, la segmentazione semantica è particolarmente impegnativa nell'architettura storica, a causa della complessità delle forme e della limitata ripetibilità di elementi in edifici diversi, rendendo difficile definire modelli comuni all'interno della stessa classe di elementi (Malinverni et al., 2019; Pierdicca et al., 2020).

Anche se la forma è ripetibile, (e.g. per un dato stile architettonico), gli oggetti sono comunque unici in quanto realizzati artigianalmente e non serializzati.

Inoltre, poiché i modelli di deep learning richiedono una quantità considerevolmente grande di dati annotati per essere addestrati al fine di gestire correttamente le scene non viste, la mancanza di (grandi) nuvole di punti annotate pubblicamente disponibili è un enorme problema, che di fatto rallenta la ricerca in questa direzione. D'altra parte, la creazione di una massa critica di nuvole di punti annotate mediante annotazione manuale è molto dispendiosa in termini di tempo e poco pratica. Per affrontare questo problema, in questo lavoro viene esplorata l'idea di sfruttare i dati sintetici delle nuvole di punti per addestrare modelli di deep learning dello stato dell'arte al fine di eseguire la segmentazione semantica di nuvole di punti TLS reali. Stante la letteratura di riferimento nel settore, questo è il primo tentativo in questa direzione.

¹ <http://buildingparser.stanford.edu/dataset.html>

Deriviamo nuvole di punti sintetiche annotate da modelli 3D esistenti con l'obiettivo di ottenere una massa critica di dati da cui apprendere, senza la necessità di annotazioni manuali dispendiose in termini di tempo, di enormi quantità di dati TLS. Sebbene le nuvole di punti sintetiche siano relativamente facili da ottenere, sono intrinsecamente diverse dai dati TLS reali, poiché hanno una struttura più regolare, non considerano gli errori accidentali degli strumenti topografici e, in genere, catturano un livello di dettaglio inferiore.

Pertanto, lo scopo della presente ricerca è fornire una prima valutazione dell'uso di dati sintetici per guidare la segmentazione semantica basata sul deep learning nel contesto degli edifici storici.

I principali contributi di questo articolo sono i seguenti:

1. Fornire il primo studio per valutare l'uso di dati sintetici di nuvole di punti nel contesto della segmentazione semantica del patrimonio costruito, valutando modelli DGCNN addestrati su dati TLS reali.
2. Rilasciare un dataset composto da nuvole di punti sintetiche annotate derivate da modelli 3D di 10 diversi edifici storici e da 159 modelli 3D di singoli elementi architettonici raccolti da archivi web aperti. Il dataset² è disponibile su richiesta agli autori e può essere utilizzato per ulteriori ricerche in questo campo.
3. Proporre una variazione dell'architettura DGCNN presentata in (Wang et al., 2019b), basata sull'uso della *radius distance*, che ha dimostrato di aumentare le prestazioni sui nostri dataset di test.

Il contributo è organizzato come segue: la Sezione 2 fornisce una panoramica del lavoro precedente riguardante i recenti sforzi nella segmentazione semantica delle nuvole di punti basata su deep learning e la generazione e l'utilizzo di dati sintetici. Nella Sezione 3 vengono descritti i dati e i metodi usati in questo lavoro. Nelle Sezioni 4 e 5 sono presentati rispettivamente gli esperimenti condotti e i loro risultati. Infine, nella Sezione 6 sono tratte le conclusioni.

2. STATO DELL'ARTE

In letteratura, l'uso di dataset sintetici coinvolge diversi campi di applicazione. Il primo lavoro risale al 1975 quando Miller (Miller and Melton, 1975) generò dati di test appropriati per ottenere il miglioramento della qualità del software a breve termine utilizzando una metodologia che mette in relazione i dati di test del software funzionale con le sue specifiche formali per ottenere una corrispondenza tra il software e le sue specifiche.

Myers (1999) ha mostrato che l'introduzione del dataset simulato porta diversi vantaggi. Ha sviluppato un software per lo studio del genoma umano e un simulatore (Celsim) che permette di descrivere e generare stocasticamente una sequenza di DNA bersaglio con diverse strutture ripetute. La generazione di dati sintetici multidimensionali è stata applicata anche al clustering e all'analisi dei valori anomali in (Pei and Zaiane, 2006).

Le immagini termiche sintetiche vengono utilizzate per aumentare un dataset esistente di immagini visibili nel lavoro di Kniaz et al. (2017). Per aumentare il dataset esistente con immagini termiche sintetiche, hanno sviluppato una rete neurale convoluzionale profonda, denominata Thermalnet, ispirata alla colorizzazione delle reti neurali profonde.

Partendo da una singola immagine bidimensionale, Fan et al. (2017) hanno ricostruito la geometria 3D dell'oggetto completo. Hanno usato un algoritmo di deep learning per generare la

²<http://vrai.dii.univpm.it/content/synthetic-dataset-point-cloud-semantic-segmentation>

rappresentazione della nuvola di punti 3D, prevedendo la forma 3D reale dell'oggetto.

Li et al. (2016) hanno proposto un approccio basato sull'apprendimento supervisionato per classificare scene urbane reali 3D acquisite da Terrestrial, Vehicle e Aerial Laser Scanner TLS, VLS e ALS. La fase di addestramento ha utilizzato un dataset di generazione automatica di nuvole di punti 3D. In (Ros et al., 2016) un dataset di immagini sintetiche di scene urbane, estratto automaticamente da scene urbane virtuali, è stato presentato e mostrato per migliorare le prestazioni di segmentazione semantica 2D e aggiunto al dataset di benchmark annotato manualmente durante la fase di addestramento. Nel lavoro di (Wu et al., 2019), nuvole di punti sintetiche realistiche di scene stradali sono derivate da un popolare videogioco (Grand Theft Auto) e utilizzate per aumentare il dataset di riferimento KITTI (Geiger et al., 2012), migliorando le prestazioni della segmentazione semantica basata sulla rete neurale convoluzionale CNN (Convolutional Neural Network).

In Wang et al. (2019a) sensori LiDAR virtuali sono stati utilizzati per acquisire dati sintetici di nuvole di punti da scene di strade, simulando una varietà di strumentazione per l'acquisizione di nuvole di punti. Gli esperimenti sulla segmentazione semantica nel contesto della guida autonoma hanno dimostrato che le prestazioni aumentano grazie all'aumento sintetico dei dati annotati manualmente.

Secondo gli autori (Jaderberg et al., 2014) un dataset sintetico di testo generato usando un motore è una soluzione ottimale per sostituire i dati reali, fornendo un'enorme quantità di dati usati per l'addestramento. Essi hanno usato un dataset sintetico per addestrare una CNN, e al fine di testare la validità dell'approccio, hanno utilizzato un dataset disponibile pubblicamente. Simile è il lavoro di Wang et al. (2012) che ha addestrato una CNN usando esempi di testo generati sinteticamente.

Il lavoro di (Griffiths and Boehm, 2019) ha utilizzato un approccio basato sull'apprendimento profondo per la classificazione automatica dei dati delle nuvole di punti 3D. A causa della necessità di una grande quantità di dati di addestramento, gli autori hanno generato SynthCity, un dataset sintetico di nuvole di punti in un ambiente urbano/suburbano utilizzando il plugin Blesor per Blender. Il loro scopo è dimostrare che una rete addestrata utilizzando dati sintetici è in grado di generalizzare bene.

Nell'ambito del patrimonio culturale i dataset sintetici non sono spesso utilizzati; tuttavia, vengono qui presentati alcuni lavori significativi (Pierdicca et al., 2019).

Martínek et al. (2019) mostra una panoramica degli approcci di addestramento per il riconoscimento ottico dei caratteri nei documenti storici. Poiché il problema principale è la mancanza di dati annotati, riassumono diversi modi per preparare i dati sintetici. Hanno addestrato un classificatore di rete neurale ricorrente convoluzionale utilizzando un dataset sintetico e convalidato il loro approccio con un dataset reale annotato.

In un altro lavoro si è utilizzato un grande dataset renderizzato sinteticamente che presenta scene con forme generate e reali per addestrare Relight-Net, una profonda rete neurale convoluzionale di (Xu et al., 2018).

Un approccio che ha utilizzato un dataset sintetico è quello proposto da (Tran and Khoshelham, 2019), dove viene presentato un confronto tra un modello di informazioni edilizie e una nuvola di punti di un ambiente interno. Usano dataset sintetici e dataset di riferimento ISPRS per valutare le prestazioni nel rilevamento delle differenze tra le due rappresentazioni.

Sebbene lo scopo di questo lavoro non è quello di fornire una revisione della letteratura sugli approcci deep learning per la segmentazione semantica delle nuvole di punti dense, vale la

pena citare alcune architetture proposte di recente. I lettori interessati possono trovare una discussione più ampia sull'argomento in (Griffith and Boehm, 2019).

Le reti neurali profonde dello stato dell'arte sono progettate specificamente per affrontare l'irregolarità delle nuvole di punti, gestendo direttamente i dati grezzi delle nuvole di punti, piuttosto che passare a una rappresentazione regolare intermedia.

Tra gli altri, PointNet (Qi et al., 2017a) è il pioniere di questo approccio. Questa rete ottiene l'invarianza di permutazione dei punti operando indipendentemente su ogni punto, e successivamente applicando una funzione simmetrica per riunire le caratteristiche. L'estensione di PointNet (Qi et al., 2017b) tiene conto dei punti vicini, piuttosto che agire su ciascuno separatamente. Questo consente lo sfruttamento delle caratteristiche locali, migliorando così le prestazioni del modello di base.

PCNN (Atzmon et al., 2018) è un framework di apprendimento profondo per l'applicazione della CNN alle nuvole di punti. La convoluzione della nuvola di punti è definita da un pull-back della convoluzione volumetrica euclidea tramite un meccanismo di restrizione di estensione.

DGCNN, Dynamic Graph Convolutional Neural Networks (Wang et al. 2019b), introduce l'operazione EdgeConv. EdgeConv è un modulo che descrive le relazioni tra un punto e i suoi vicini. Questo modulo è progettato per essere invariante all'ordinamento dei vicini ed alle permutazioni. Nei nostri esperimenti ci basiamo su DGCNN e valutiamo i modelli addestrati su due nuvole di punti TLS acquisite da due diversi edifici storici: il Palazzo Ducale di Urbino e Palazzo Ferretti in Ancona.

3. MATERIALI E METODI

3.1 Classi degli elementi architettonici

Il primo passo fondamentale per la metodologia qui illustrata è identificare le classi di elementi architettonici da considerare per il compito di segmentazione semantica. Le classi selezionate per gli esperimenti sono le seguenti: Colonna, Finestra/Porta, Muro, Pilastro, Pavimento, Modanatura, Volta, altro. Le classi e il loro livello di dettaglio (LOD) sono stati scelti tenendo conto di tre aspetti: coerenza con il dizionario esistente, coerenza con il periodo storico e il suo linguaggio e corrispondenza con i dati acquisiti tramite TLS.

Per quanto riguarda il dettaglio scelto, si è fatto riferimento alle suddivisioni ontologiche proposte da robusti dizionari, disponibili in rete, relativi al campo dell'arte e dell'architettura. Per questo lavoro il riferimento principale è il vocabolario degli elementi architettonici classici e le loro gerarchie sviluppate dal Getty Museum – Research³. Il livello gerarchico della divisione in classi è stabilito tra elementi strutturali e componenti decorativi. Il grado di degrado dell'edificio consente di avere una piena corrispondenza con questo primo test condotto. Infatti, l'accuratezza attesa della nuvola di punti, in accordo con la letteratura di riferimento nonostante sia utilizzata in altri domini, è conforme al livello generale di dettaglio degli elementi architettonici.

Le classi scelte, inoltre, sono coerenti con i principali elementi del linguaggio architettonico classico, massicciamente utilizzato negli edifici rinascimentali (presi come casi studio di riferimento).

Per alcune classi semantiche è stato necessario implementare il dataset disponibile con alcuni modelli 3D online.

³<http://www.getty.edu/research/tools/vocabularies/aat/>.

Le classi architettoniche sono inoltre conformi ai dati acquisiti tramite TLS.

In relazione a questa metodologia di indagine e al fine di mantenere la massima uniformità ai dati acquisiti, alcune classi (come il tetto) non sono state prese in considerazione pur essendo nei vocabolari e nel linguaggio storico architettonico.

Come nel caso di qualsiasi architettura storica di grande valore e complessità, nonostante l'ampiezza delle classi, alcuni elementi non sono stati inclusi in nessuna di queste. Tutti gli elementi architettonici che non corrispondevano a nessuna delle classi selezionate sono stati raggruppati nella classe *Altro*.

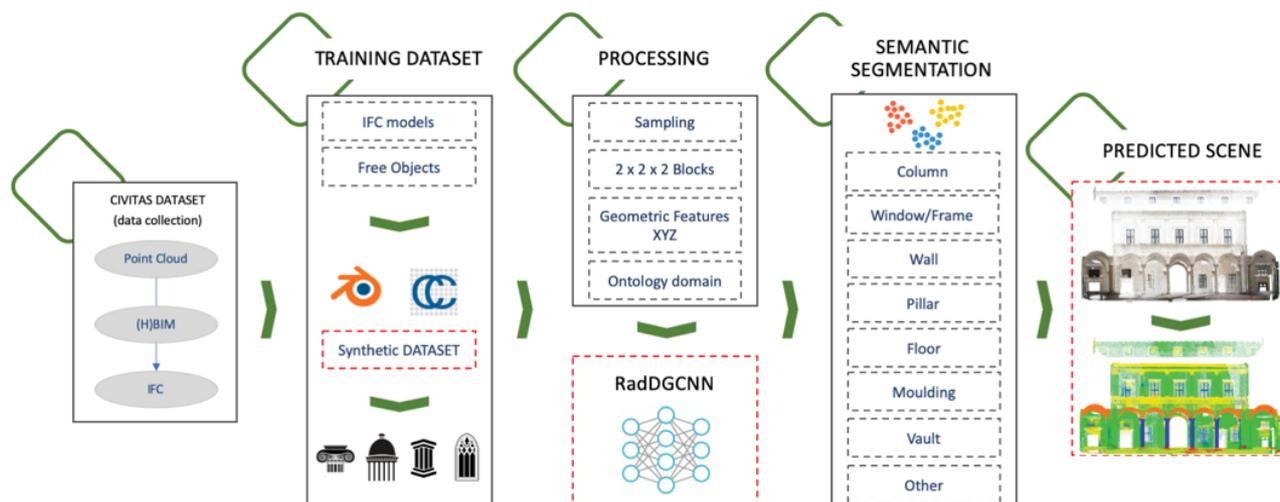


Figura 1. Workflow della segmentazione semantica proposta

3.2 Dataset di addestramento sintetico

In questa sezione viene descritto il dataset utilizzato nei nostri esperimenti per addestrare il modello neurale profondo. Il dataset è costituito da nuvole di punti derivate da modelli 3D esistenti ed è composto da due parti.

Il primo include nuvole di punti estratte da modelli 3D di edifici storici reali. Nella sezione 3.2.1 sono forniti dettagli sul processo di modellazione *Scan-to-BIM* svolto dagli esperti come parte della loro attività di ricerca.

La seconda parte include nuvole di punti derivate da una raccolta di oggetti 3D di singoli elementi architettonici, raccolti da repository online aperti, descritti in dettaglio nella sezione 3.2.2.

In entrambi i casi, è stato utilizzato il software open source CloudCompare⁴ per convertire i modelli 3D, in formato OBJ, in nuvole di punti, campionando casualmente punti da superfici con una densità di 1000 punti per m². Lo stesso sottocampionamento è stato applicato alle nuvole di punti TLS utilizzate per testare le reti neurali.

3.2.1 Modelli 3D dai tradizionali flussi di lavoro Scan-to-BIM: la sfida principale nel campo HBIM è incorporare l'enorme conoscenza dell'architettura storica disponibile da vari domini in modelli affidabili e accurati, sfruttando le piattaforme esistenti e le loro potenzialità in intelligenza semantica, informazioni parametriche e alimentazione di glossari. L'approccio proposto in questo articolo si basa sulla precedente strutturazione semantica e standardizzazione di un insieme significativo di edifici, grazie al solido know-how del gruppo di ricerca sulla digitalizzazione dell'architettura storica e sulla sua rappresentazione (Quattrini et al., 2017).

Un efficace approccio BIM per la documentazione del patrimonio architettonico (Apollonio et al., 2012) è stato applicato negli ultimi anni per lo sviluppo di diversi casi studio (Quattrini et al., 2016), dimostrando chiaramente una

sostanziale coerenza tra il modo classico di costruire e la possibilità di rappresentarlo in modelli semantici (Bianconi et al., 2018; Quattrini et al., 2017). Un interessante approccio metodologico mostra una descrizione semantica degli elementi architettonici basata su riflessioni teoriche ed esperienze di ricerca. Un concetto comune e convalidato è quello di estrarre la conoscenza dell'architettura classica da trattati così come da indagini accurate e, quindi, sviluppare modelli parametrici a diversi livelli di oggetti (Bianchini et al., 2017). Le varianti derivano dai modelli parametrici modificando o raggruppando i parametri in base alle grammatiche o alla consapevolezza degli operatori.

Il primo approccio all'HBIM nel Palazzo Ducale è stato una fase di modellazione di base e già valutata, basata su strumenti di editing di Revit come le famiglie caricabili o locali. Tuttavia, la parametrizzazione e l'adozione di robuste tassonomie sono intese come caratteristiche fondamentali per modelli informati. Al fine di effettuare un'analisi approfondita e sviluppare tutte le parti in modo rigoroso, è stato modellato un primo nucleo del Cortile d'Onore (Figura 2). La scansione ritmica e la modularità dello spazio progettato da Luciano Laurana racchiudono tutti gli elementi in grado di sintetizzare l'intero Palazzo. Gli schemi degli elementi architettonici sono stati realizzati partendo dall'integrazione dei dati di rilievo con la letteratura classica e con riferimenti più recenti, ottenendo anche un riferimento per affrontare il tema dei diversi LOD (level of detail, livello di sviluppo e definizione).

Per quanto riguarda le strutture verticali, l'elemento principale è un ordine corinzio con suddivisione che attualmente non è stato considerato nel raggruppamento successivo. Mentre tutte le aperture erano distribuite semanticamente in "finestra" e "cornici". Per quanto riguarda le famiglie caricabili "colonna", "finestra" e "pilastrini", gli oggetti sono stati modellati impostando vincoli geometrici in grado di sviluppare una parametrizzazione. Gli angoli del Cortile necessitavano di famiglie locali "ad hoc", secondo il disegno innovativo di una "soluzione d'angolo" ben nota in storiografia e letteratura.

⁴<https://www.danielgm.net/cc/>

Un altro passo importante nella modellazione, inoculando funzionalità ricche nel presente framework, è la definizione di LOD, che consente diversi livelli di rappresentazione della realtà, nonché di estrarre nuvole di punti modellate secondo dettagli corretti. Il modello 3D ottenuto viene informato, annotato e dettagliato.

Nel caso di Palazzo Ferretti (Figura 3), la procedura aveva previsto di costruire un modello parametrico dell'edificio in un ambiente Revit integrando più dati: rilievo TLS, rilievi bidimensionali e rilievo fotogrammetrico per informazioni di dettaglio, nonché un confronto continuo con le regole dei trattati. Il lavoro principale riguarda le modanature delle facciate principali con la realizzazione di famiglie caricabili (*.rfa) incorporanti semantiche e vincoli geometrici in coerenza con la pratica dell'architettura classica (Moreira et al., 2018). I risultati più significativi sinora per Palazzo Ferretti riguardano la realizzazione di librerie parametriche per le modanature esterne ed interne degli edifici, ora quelle modanature possono alimentare il nostro dataset sintetico.



Figura 2. Palazzo Ducale di Urbino, il Cortile d'Onore progettato da Luciano Laurana



Figura 3. Palazzo Ferretti, facciata principale

Considerando la necessità di ampliare il dataset semanticamente strutturato, sono stati sfruttati i modelli provenienti dalla Palladio Library (ricerca di rilievo e documentazione digitale delle ville e dei ponti di Andrea Palladio nell'ambito del progetto finanziato da Arcus/CISA Andrea Palladio, per la realizzazione del progetto "Biblioteca Palladio - mediateca e sistema di comunicazione e informazione su Andrea Palladio, le sue opere e il suo territorio", responsabile scientifico dell'Unità UNIVPM: Paolo Clini, responsabile scientifico dell'intero

progetto: Marco Gaiani). Dall'idea di Biblioteca Palladiana, è stato principalmente prodotto un geo-database (Apollonio et al., 2013; Gaiani et al., 2015) che costituisce una sorta di HBIM ante litteram.

Il gruppo di modelli sviluppati per la Biblioteca Palladiana comprendeva 13 modelli, in particolare due ponti e varie ville e/o barchesse. Per il presente lavoro sono state prese in considerazione solo le ville perché presentavano classi coerenti con quelle delle nuvole di punti che dovevano essere segmentate. Una caratteristica principale di questi modelli, alla luce del presente lavoro, è stata la *shape-grammar* adottata nella costruzione dei modelli 3D utilizzando un insieme prestabilito di regole formali a forma di albero che indicano uno scopo chiaro e una struttura evidente. Questa organizzazione è stata estesa, nei nostri casi, su più livelli gerarchici, consentendo ora di scoprire la corrispondenza con il clustering precedente. I modelli della Palladio Library sono stati infatti archiviati in formato file 3DM incorporando una tassonomia grazie alla denominazione dei layer, consentendo l'esportazione in formato OBJ e la successiva creazione di nuvole di punti sintetiche strutturate semanticamente.

Da tutti i modelli sopra menzionati, progettati da esperti di dominio, è stato possibile ottenere il nucleo del dataset sintetico sufficientemente robusto da istanziare il flusso di lavoro successivo.

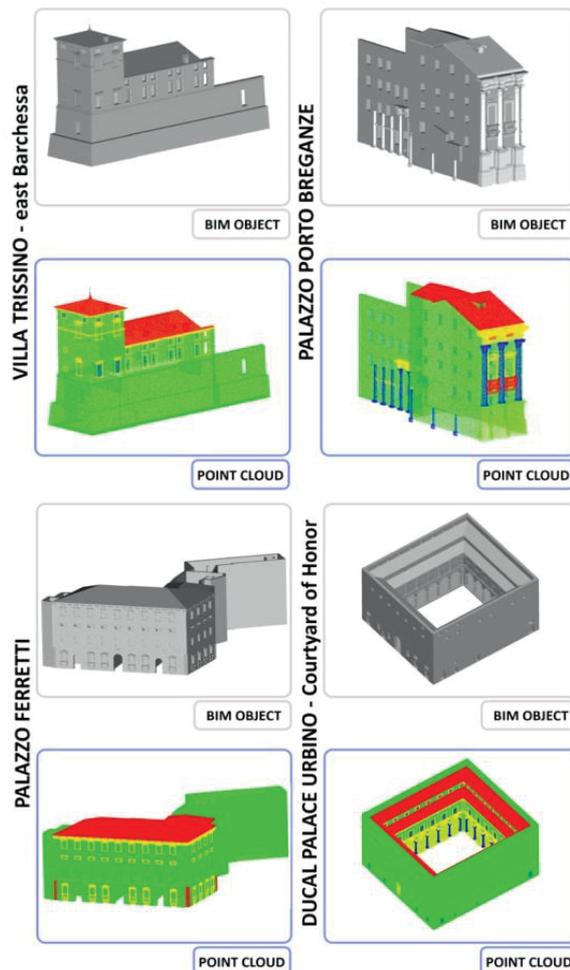


Figura 4. Esempi di modelli 3D di edifici storici e nuvole di punti annotate derivate

3.2.2 Oggetti 3D dagli archivi online: mentre alcune classi, e.g. parete e pavimento, sono ampiamente presenti nei modelli 3D sopra descritti, lo stesso non vale per altre classi di elementi architettonici. Ad esempio, non tutti gli edifici hanno colonne o volte e, quando questi elementi sono presenti, sono solitamente molto simili all'interno dello stesso edificio. Al contrario, per apprendere caratteristiche generalizzabili, dobbiamo migliorare la variabilità della forma all'interno della stessa classe.

Nel tentativo di affrontare tali problemi, abbiamo esteso il nostro dataset sintetico con modelli 3D di singoli elementi architettonici, selezionando oggetti 3D disponibili pubblicamente da una serie di repository online. (<http://www.cgtrader.com>; <http://www.3dcadbrower.com>; <https://www.turbosquid.com>; <https://free3d.com>; <https://archive3d.net>; <https://3dwarehouse.sketchup.com>).

Nella Tabella 1 riportiamo, per ogni classe, il numero di modelli 3D raccolti, che possono includere uno o più elementi architettonici, e il numero totale di punti ottenuti trasformando tali modelli in nuvole di punti.

Classe	# Modelli 3d	# Punti
Colonne	53	448.297
Finestre/porte	41	1.041.552
Modanature	31	442.809
Pilastrini	24	558.380
Volte	10	3.743.896

Tabella 1. Panoramica degli oggetti 3D raccolti dai repository online

Come mostrato in Figura 5 i tipi di oggetti sono stati scelti come un affinamento delle classi che mancano, dando priorità a modelli di complessità e qualità paragonabili a quelli modellati dagli autori.

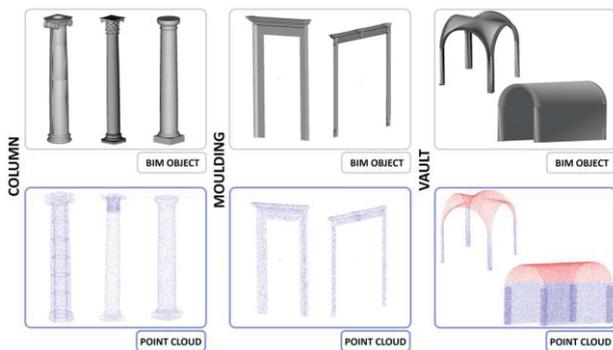


Figura 5. Esempi di modelli 3D di singoli elementi architettonici di classi diverse e corrispondenti a nuvole di punti sintetiche

3.3 Rilevamento delle nuvole di punti

Come spiegato in (Clini et al., 2018, Nespeca, 2019) la digitalizzazione tridimensionale del Palazzo Ducale di Urbino è un primo risultato consentito dal Progetto Strategico CIVITAS. In questo contesto è stata condotta una campagna di indagine completa, basata principalmente su TLS e acquisizione di dati fotografici. Allo stato attuale, l'intero modello numerico del Palazzo è costituito da 1.790 mln di punti, mentre l'esperimento si occupa di una parte della nuvola di punti relativa al Cortile d'Onore (Tabella 2).

La prima nuvola di punti concepita come cornice è un modello completo per questo edificio di grandi dimensioni: un principale dataset di riferimento è stato fornito dall'acquisizione del Leica Backpack. Per il lavoro in corso, invece, si preferisce utilizzare

la nuvola di punti realizzata grazie al TLS (Leica ScanStation C10 e Leica ScanStation P40) montato su treppiede. Questa acquisizione è stata eseguita impostando diversi livelli di risoluzione in base alla complessità degli ambienti e dei loro elementi decorativi, ottimizzando i tempi di scansione. Altre acquisizioni sono attualmente in corso con KAARTA Stencil V2 e UAV per completare il modello numerico.

Palazzo Ducale di Urbino		
	Nuvola intera	Nuvola annotata
Numero di punti	1.224 mln	17,3 mln
Errore max di align.	0,006 m	0,008 m
Errore min di align.	0,001 m	0,003 m
RMS	0,005 m	0,004 m
Palazzo Ferretti		
	Nuvola intera	Nuvola annotata
Numero di punti	1.273 mln	8,4 mln
Errore max di align	0,008	0,005 m
Errore min di align	0,003	0,001m
RMS	0,007	0,007 m

Tabella 2. Numero di punti, errore di allineamento max / min e RMS delle due nuvole di punti TLS utilizzate nei nostri esperimenti

Le operazioni di rilievo TLS del complesso di Palazzo Ferretti risalgono al 2014: sono state effettuate 69 stazioni: 49 stazioni (58 scansioni) per il rilievo esterno dell'edificio e 20 stazioni e scansioni per il rilievo interno, del piano terra e del piano primo. Queste scansioni sono state acquisite con una risoluzione di 1 cm a 100 m, alcune parti in cui erano necessari maggiori dettagli sono state acquisite con una nuvola di punti più densa (0,5 cm a 100 m). La nuvola di punti finale ottenuta dall'allineamento delle nuvole di tutte le 78 scansioni è composta da 1,2 miliardi di punti (vedi Tabella 2). Per questa ricerca una parte della nuvola di punti, riferita all'età rinascimentale, è stata sfruttata per coprire le classi già disponibili nel Palazzo Ducale di Urbino. Solo una parte dell'intera nuvola di punti è stata quindi annotata (Tabella 2).

3.4 Modello di rete neurale profonda

Dynamic Graph Convolutional Neural Networks (DGCNN), recentemente proposta in (Wang et al., 2019b), ha dimostrato di fornire prestazioni di segmentazione semantica su benchmark standard dello stato dell'arte, come il dataset S3DIS (Armeni et al., 2016).

In questo documento abbiamo considerato DGCNN e una sua nuova versione modificata (RadDGCNN) introdotta in questo lavoro; successivamente abbiamo combinato i due approcci per aumentare ulteriormente le prestazioni.

3.4.1 Architettura: il modello DGCNN si basa sull'operazione EdgeConv, introdotta in (Wang et al., 2019b), che apprende le caratteristiche nascoste di un punto in base ai punti vicini. Nonostante il suo nome, EdgeConv è implementato con un Multilayer Perceptron (MLP) alimentato con le cosiddette *edge features* e da una successiva operazione di max-pooling sulle caratteristiche apprese. Nel nostro caso, le caratteristiche del bordo di un punto sono composte dal vettore tridimensionale che rappresenta le coordinate del punto e i vettori di distanza tra il punto e i punti vicini.

L'operazione EdgeConv viene inizialmente eseguita sulla nuvola di punti di input e quindi, ripetutamente, sull'output del livello EdgeConv precedente, apprendendo così le caratteristiche nascoste locali gerarchiche per ogni punto.

Le funzionalità nascoste di output di tutti i punti appresi dai livelli EdgeConv vengono concatenate e inviate a una rete MLP per apprendere le funzionalità globali della nuvola di punti.

Infine, le caratteristiche globali e le caratteristiche locali vengono concatenate e inviate ad un altro MLP che fornisce la categoria prevista per ogni punto.

L'architettura di rete utilizzata nei nostri esperimenti è rappresentata nella Figura 6. In questa figura, N rappresenta il numero di punti per ogni nuvola di punti di input e P è il

numero di classi considerate per la classificazione basata su punti e K è il numero di punti vicini considerati per ogni punto. Tre livelli EdgeConv vengono utilizzati per apprendere le caratteristiche del punto locale. Le funzionalità apprese da ogni livello EdgeConv per ogni punto vengono quindi concatenate e le funzionalità globali vengono apprese con un MLP a tre livelli.

Abbiamo adottato le stesse impostazioni, dimensioni di MLP e numero di livelli, utilizzati in (Wang et al., 2019b).

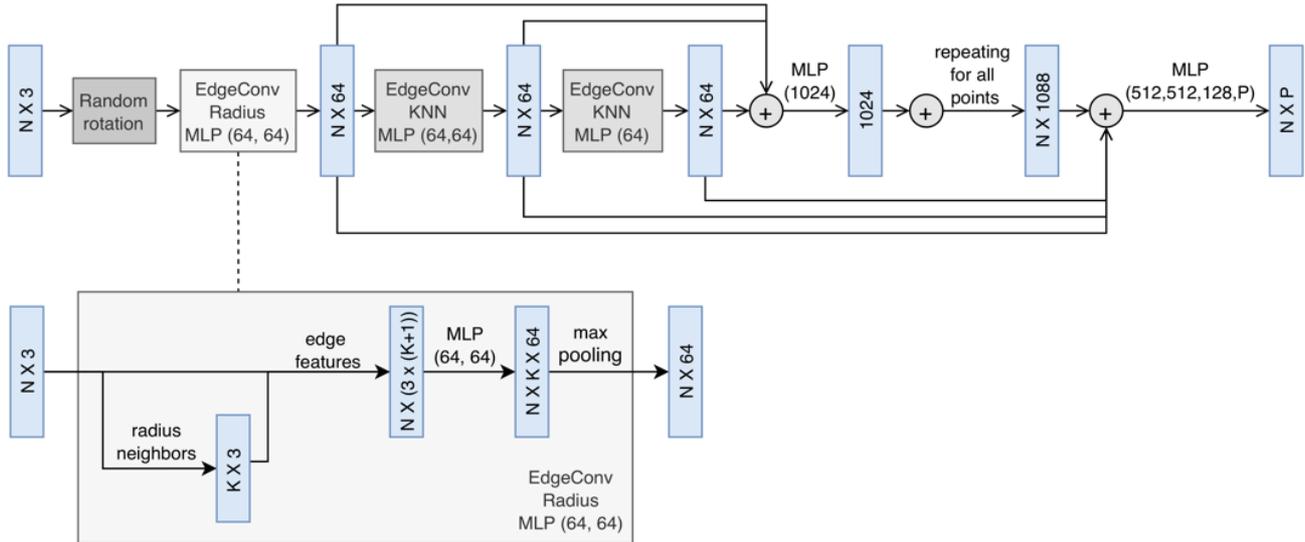


Figura 6. Architettura della rete neurale profonda adottata in fase sperimentale

3.4.2 Vicinanza dei punti: Una decisione importante da prendere nell'attuazione di una DGCNN è come definire la vicinanza di un punto. In (Wang et al., 2019b), i K-Nearest Neighbors (KNN) di un punto sono considerati per calcolare le caratteristiche del bordo ad ogni successiva operazione EdgeConv. Tuttavia, come suggerito da recenti studi (Hermosilla et al., 2018), KNN potrebbe non essere il metodo migliore quando si tratta di nuvole di punti non uniformemente dense. Questo è proprio il nostro caso, poiché l'acquisizione laser produce generalmente nuvole di punti non uniformi. Nel caso dei KNN, l'estensione spaziale dei vicini considerati per i due punti può essere molto diversa, a seconda della densità locale della nuvola di punti. Questo a sua volta, potrebbe portare a caratteristiche di bordo diverse per due punti che appartengono alla stessa classe (finestra). D'altra parte, usando la distanza del raggio, i punti vicini coprono la stessa area.

Spinti da questa considerazione, abbiamo sperimentato una variazione dell'architettura DGCNN (RadDGCNN in questo articolo), dove la distanza del raggio viene utilizzata nella prima operazione EdgeConv al posto del KNN (Figura 4)).

In pratica, per ogni punto, si selezionano prima tutti i punti entro la distanza raggio R dal punto stesso, quindi si campionano casualmente K di questi punti e li forniamo come input all'MLP per estrarre le caratteristiche del bordo. Tale operazione di campionamento ci consente di fornire sempre lo stesso numero di punti vicini (e quindi lo stesso numero di caratteristiche di fronte) come input per il MLP.

4. SETUP SPERIMENTALE

Abbiamo addestrato i modelli DGCNN e RadDGCNN sul nostro dataset sintetico suddividendo ciascuna nuvola di punti in blocchi di dimensioni uguali (2 metri x 2 metri x 2 metri)

ciascuno contenente 4096 punti campionati casualmente. Abbiamo fatto lo stesso per il set di prova.

La segmentazione della nuvola di punti, utilizzata anche in lavori precedenti sulla segmentazione della scena tramite reti neurali, è una fase di preelaborazione obbligatoria poiché fornire le intere nuvole di punti come input sarebbe computazionalmente irrealizzabile. Il campionamento casuale dei punti per ogni blocco ci permette di ottenere input di uguali dimensioni, che è un requisito dell'architettura DGCNN.

Nei nostri esperimenti, ogni punto è rappresentato da un vettore di 6 dimensioni, dove i primi 3 elementi sono le coordinate XYZ assolute del punto nella nuvola di punti originale, mentre gli altri 3 valori rappresentano le coordinate normalizzate all'interno del blocco (nell'intervallo [0-1]).

Per consentire un confronto equo, abbiamo adottato gli stessi iperparametri di rete di (Wang et al, 2019b), con l'unica differenza nella dimensione dei mini-batch, che abbiamo impostato a 16. Più specificamente, il tasso di apprendimento iniziale è stato impostato a 0,001 ed è stato utilizzato un decadimento della velocità di apprendimento di 0,5; lo slancio per la normalizzazione del batch è stato impostato su 0,9; e abbiamo utilizzato l'algoritmo di ottimizzazione di Adam per la retro-propagazione. Infine, è stata applicata una regolarizzazione di abbandono con probabilità di mantenimento di 0,5 negli ultimi due strati completamente connessi.

Per quanto riguarda l'approccio RadDGCNN, la distanza del raggio è stata calcolata sulle coordinate normalizzate, R è stato impostato sperimentalmente a 0,1. In entrambi i modelli il valore di K è stato impostato a 20, seguendo le indicazioni fornite in (Wang et al., 2019b).

A differenza di altri studi, non abbiamo utilizzato le caratteristiche del colore (es. RGB o HSV) poiché il dataset

sintetico è derivato da modelli 3D, dove i colori di solito non corrispondono al colore reale degli elementi architettonici.

Infine, prima di inserire un blocco nella rete, i punti sono stati ruotati casualmente attorno alla direzione verso l'alto. In questo modo viene incrementato il dataset per tenere conto del fatto che gli elementi architettonici possono avere un orientamento diverso. Sebbene ulteriori tecniche di potenziamento dei dati potrebbero essere applicate per aumentare eventualmente la capacità della rete di apprendere forme complesse, adottando approcci come quello descritto in (Pierdicca et al., 2020), questo viene rimandato ad indagini future.

In questa ricerca sono stati eseguiti due esperimenti su due diversi casi di studio.

Nel primo è stata usata la nuvola di punti TLS del cortile del Palazzo Ducale di Urbino per valutare i modelli addestrati, cercando di individuare 8 classi di elementi architettonici differenti. Per affrontare uno scenario del mondo reale, è stata rimossa dal dataset la nuvola di punti sintetica derivata dal modello BIM del Palazzo Ducale di Urbino.

Nel secondo esperimento sono stati valutati i modelli addestrati sulla nuvola di punti TLS di Palazzo Ferretti. In questo caso dal dataset addestrato è stato escluso il modello BIM di Palazzo Ferretti. Poiché due delle classi di elementi architettonici selezionati (colonna e pilastro) non sono presenti nell'edificio, si è cercato di riconoscere le restanti 6 classi.

Infine, in entrambi gli esperimenti i due modelli sono stati combinati implementando un semplice sistema multi-

classificatore. Per ogni punto nelle nuvole TLS di test, è stato calcolato il risultato di probabilità assegnato a ciascuna classe dal DGCNN e dal RadDGCNN, selezionando la classe con il punteggio più alto.

5. RISULTATI E DISCUSSIONE

Nella Tabella 3 vengono riportati i risultati del primo esperimento, ottenuto sulla scena del test (Cortile del Palazzo Ducale di Urbino) utilizzando DGCNN, RadDGCNN e l'approccio combinato. I risultati del secondo esperimento, valutati su Palazzo Ferretti, sono riportati nella Tabella 4.

Per ogni classe si hanno precisione, recall, F1score e Intersection over Union (IoU). In entrambe le tabelle, i migliori risultati sono in grassetto, mentre i valori sottolineati indicano i migliori risultati tra DGCNN e RadDGCNN.

I risultati qualitativi sono illustrati nella Figura 7, nella quale sono riportate porzioni delle nuvole di punti dei due edifici storici, mostrando i dati TLS originali, la nuvola di punti annotata manualmente (usata come *ground truth* nella valutazione) e l'output del modello RadDGCNN addestrato (predizione).

Infine, la classe *Altro*, include un'enorme varietà di elementi diversi (es. scale, bugnati e diversi elementi decorativi), che sono chiaramente difficili da classificare automaticamente e vengono spesso confusi con muri, modanature o finestre.

classes	DGCNN [36]				RadDGCNN				Combined			
	precision	recall	f1-score	IoU	precision	recall	f1-score	IoU	precision	recall	f1-score	IoU
Column	0.8493	0.8387	0.8440	0.7301	0.9807	<u>0.8414</u>	0.9057	0.8277	0.8647	0.8958	0.8800	0.7857
Window/Door	0.2530	0.4067	0.3119	0.1848	<u>0.4950</u>	0.5864	0.5368	0.3669	0.5679	0.5540	0.5609	0.3897
Wall	0.7458	0.7057	0.7252	0.5689	<u>0.7650</u>	<u>0.8499</u>	<u>0.8052</u>	<u>0.6739</u>	0.7680	0.8590	0.8110	0.6821
Pilaster	0.2066	0.0120	0.0226	0.0114	0.0626	0.0006	0.0012	0.0006	0.0762	0.0004	0.0007	0.0004
Floor	0.9957	0.7221	0.8371	0.7198	0.9107	<u>0.9266</u>	<u>0.9186</u>	<u>0.8494</u>	0.9305	0.9414	0.9359	0.8795
Moulding	0.5057	0.5870	0.5433	0.3730	<u>0.7046</u>	<u>0.6266</u>	<u>0.6633</u>	<u>0.4962</u>	0.7173	0.6409	0.6770	0.5117
Vault	0.7094	0.9378	0.8077	0.6775	0.7820	0.8684	<u>0.8229</u>	<u>0.6991</u>	0.7780	0.8948	0.8323	0.7128
Other	0.0092	0.0074	0.0082	0.0041	<u>0.0273</u>	0.0264	<u>0.0268</u>	<u>0.0136</u>	0.0377	0.0220	0.0278	0.0141
micro avg			0.6714				<u>0.7619</u>					0.7747
macro avg	0.5343	0.5272	0.5125	0.4087	<u>0.5910</u>	<u>0.5908</u>	<u>0.5851</u>	<u>0.4909</u>	0.5925	0.6010	0.5907	0.4970
weighted avg	0.6812	0.6714	0.6650		<u>0.7322</u>	<u>0.7619</u>	<u>0.7436</u>		0.7363	0.7747	0.7518	

Tabella 3. Risultati della segmentazione semantica di 8 classi ottenute sui dati TLS da Palazzo Ducale a Urbino.

classes	DGCNN [36]				RadDGCNN				Combined			
	precision	recall	f1-score	IoU	precision	recall	f1-score	IoU	precision	recall	f1-score	IoU
Window/Door	0.3443	0.4992	0.4075	0.2559	0.4160	0.4306	<u>0.4232</u>	<u>0.2684</u>	0.4135	0.4772	0.4431	0.2846
Wall	0.8032	0.7591	0.7805	0.6400	<u>0.8276</u>	<u>0.7826</u>	<u>0.8045</u>	0.6729	0.8408	0.8304	0.8356	0.7175
Floor	0.9833	0.7344	0.8408	0.7254	<u>0.9913</u>	0.7315	<u>0.8418</u>	<u>0.7268</u>	0.9944	0.7326	0.8437	0.7296
Moulding	0.6016	0.6829	0.6397	0.4702	0.4770	<u>0.7063</u>	0.5694	0.3980	0.5558	0.7146	0.6253	0.4548
Vault	0.9282	<u>0.6348</u>	<u>0.7539</u>	0.6050	0.9145	0.6215	0.7401	0.5874	0.9171	0.6587	0.7667	0.6217
Other	<u>0.1845</u>	0.2807	0.2226	0.1253	0.1273	0.1120	0.1192	0.0634	0.2127	0.1878	0.1995	0.1108
micro avg			0.6830				<u>0.6834</u>					0.7202
macro avg	<u>0.6408</u>	<u>0.5985</u>	<u>0.6075</u>	<u>0.4703</u>	0.6256	0.5641	0.5830	0.4528	0.6557	0.6002	0.6190	0.4865
weighted avg	<u>0.7284</u>	0.6830	<u>0.6981</u>		0.7197	<u>0.6834</u>	0.6919		0.7444	0.7202	0.7259	

Tabella 4. Risultati della segmentazione semantica di 6 classi ottenuti sui dati TLS di Palazzo Ferretti.

Mentre nel primo esperimento, RadDGCNN supera sensibilmente l'approccio basato su KNN, nel secondo esperimento forniscono risultati complessivi comparabili.

In entrambi gli esperimenti, la combinazione di reti basate su KNN e Radius fornisce i migliori risultati complessivi, raggiungendo rispettivamente lo 0,77 e lo 0,72 di accuratezza complessiva nella segmentazione del Palazzo Ducale e di Palazzo Ferretti. Ciò suggerisce che le due funzioni adiacenti sono di fatto complementari e possono essere efficacemente utilizzate insieme per aumentare le prestazioni.

In entrambi i casi, i pavimenti, i muri e le volte sono relativamente ben segmentati, con F1score che vanno da 0,77 a

0,93 e IoU che vanno da 0,62 e 0,87. Anche le colonne, presenti solo nel Palazzo Ducale, vengono rilevate con precisione (0,88 F1score e 0,78 IoU).

Risultati leggermente peggiori si ottengono per finestre/porte e modanature. Questo era in qualche modo prevedibile poiché entrambe le classi presentano un'elevata variabilità di forma e quindi sono più difficili da segmentare. Infatti, poiché le modanature includono anche quelle che incorniciano finestre e porte, sono spesso confuse con finestre e porte poiché i due elementi hanno generalmente una forma simile. Questo è visibile anche esaminando le matrici di confusione, mostrate nella Figura 8.

I risultati evidenziano anche come i pilastri siano molto difficili da rilevare, in quanto quasi indistinguibili dai muri (vedi Figura 8a), se non per la presenza del capitello, che quasi sempre è classificato come modanatura.

Nel complesso, riteniamo che i risultati ottenuti dimostrano che l'uso di dati sintetici può realizzare efficacemente la segmentazione automatica delle nuvole di punti TLS e dovrebbero essere ulteriormente studiati per supportare il processo *Scan-to-BIM* degli edifici storici.



Figura 7. Visualizzazione dei risultati della segmentazione automatica di Palazzo Ducale (a) e Palazzo Ferretti (b)

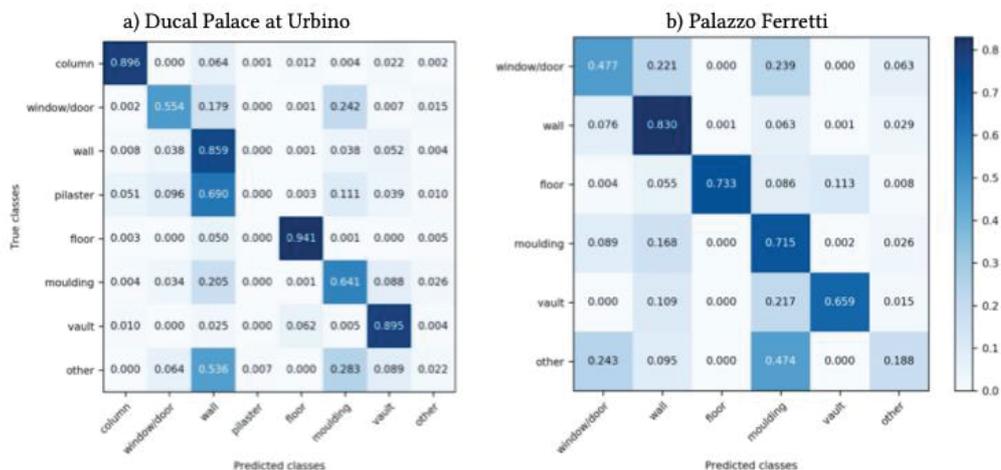


Figura 8. Matrice di confusione della segmentazione semantica dei dati TLS del Palazzo Ducale di Urbino (a) e Palazzo Ferretti (b), ottenuta con il metodo combinato.

6. CONCLUSIONI E LAVORI FUTURI

La segmentazione semantica delle nuvole di punti è un compito rilevante nel CH, in quanto facilita l'identificazione automatica di elementi architettonici storici, possibilmente accelerando il processo *Scan-to-BIM*. I recenti miglioramenti negli approcci di apprendimento profondo hanno fornito risultati incoraggianti nella segmentazione semantica delle scene interne ed esterne. Tuttavia, nel contesto degli edifici storici, questo compito è reso particolarmente impegnativo dalla complessità e dall'elevata variabilità degli oggetti da rilevare e dalla mancanza di dataset annotati disponibili. Per affrontare questa sfida di ricerca, in questo articolo abbiamo fornito il primo studio sulla valutazione dell'uso di dati sintetici per addestrare modelli di apprendimento profondo per segmentare semanticamente nuvole di punti TLS

da edifici storici, con l'obiettivo finale di aiutare ad automatizzare il *Scan-to-BIM*, che, nel dominio CH, viene eseguito principalmente manualmente.

I risultati ottenuti su due diversi casi di studio, utilizzando sia un metodo dello stato dell'arte (DGCNN) che una sua versione modificata (RadDGCNN), forniscono risultati incoraggianti, raggiungendo buoni livelli di accuratezza.

In questo studio ci siamo basati su un dataset di addestramento sintetico relativamente piccolo e riteniamo che i risultati potrebbero essere notevolmente migliorati aumentando la quantità di dati sintetici. Questo, a sua volta, fornisce potenzialmente un incentivo a condividere modelli BIM realizzati all'interno della comunità del patrimonio culturale digitale. Nei lavori futuri abbiamo in programma di mettere in atto ulteriori strategie di aumento dei dati oltre la semplice

rotazione casuale, ad es. utilizzando lo strumento presentato in (Pierdicca et al., 2019).

Un'altra preziosa linea di ricerca sarà la combinazione di dataset sintetici e reali, al fine di ampliare il set di addestramento. In tal senso, le nuove frontiere delle Reti Generative Avversarie (note in letteratura come GAN) hanno dimostrato di poter generare dati con ottimi livelli di affidabilità e similarità con il reale. Nonostante le note complessità ed unicità degli elementi architettonici, i modelli di apprendimento generativi possono risultare utili per generare dati sintetici che poi potranno essere sfruttati sia per facilitare l'addestramento delle reti profonde, sia per ottimizzare il processo di Scan-to-Bim.

6.1 Bibliografia

- Apollonio, F.I., Baldissini, S., Clini, P., Gaiani, M., Palestini, C. and Trevisan, C., 2013. The PALLADIOLibrary geo-models: An open 3D archive to manage and visualize information-communication resources about Palladio. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 5, p.W2.
- Apollonio, F.I., Gaiani, M. and Sun, Z., 2012. BIM-based modeling and data enrichment of classical architectural buildings. *SCIRES-IT-SCientific REsearch and Information Technology*, 2(2), pp.41-62.
- Armeni, I., Sener, O., Zamir, A.R., Jiang, H., Brilakis, I., Fischer, M. and Savarese, S., 2016. 3d semantic parsing of large-scale indoor spaces. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1534-1543).
- Atzmon, M., Maron, H. and Lipman, Y., 2018. Point convolutional neural networks by extension operators. *arXiv preprint arXiv:1803.10091*.
- Bianchini, C., Inglese, C., Ippolito, A., Maiorino, D. and Senatore, L.J., 2017. Building Information Modeling (BIM): Great Misunderstanding or Potential Opportunities for the Design Disciplines?. In *Handbook of Research on Emerging Technologies for Digital Preservation and Information Modeling* (pp. 67-90). IGI Global.
- Bianconi, F., Filippucci, M. and Meconi, F.M., 2018. Parametrical Vitruvius. generative modeling of the architectural orders. *SCIRES-IT*, 8(2).
- Clini, P., Quattrini, R., Bonvini, P., Nespeca, R., Angeloni, R., Mammoli, R., Dragoni, A.F., Morbidoni, C., Sernani, P., Mengoni, M. and Leopardi, A., 2020. Digit (al) isation in Museums: Civitas Project—AR, VR, Multisensorial and Multiuser Experiences at the Urbino's Ducal Palace. In *Virtual and Augmented Reality in Education, Art, and Museums* (pp. 194-228). IGI Global.
- Fan, H., Su, H. and Guibas, L.J., 2017. A point set generation network for 3d object reconstruction from a single image. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 605-613).
- Gaiani, M., Apollonio, F.I., Clini, P. and Quattrini, R., 2015. A mono-instrumental approach to high-quality 3D reality-based semantic models application on the PALLADIO library. In *2015 Digital Heritage* (Vol. 2, pp. 29-36). IEEE.
- Geiger, A., Lenz, P. and Urtasun, R., 2012, June. Are we ready for autonomous driving? the kitti vision benchmark suite. In *2012 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 3354-3361). IEEE.
- Griffiths, D. and Boehm, J., 2019. SynthCity: A large scale synthetic point cloud. *arXiv preprint arXiv:1907.04758*.
- Hermosilla, P., Ritschel, T., Vázquez, P.P., Vinacua, À. and Ropinski, T., 2018. Monte carlo convolution for learning on non-uniformly sampled point clouds. *ACM Transactions on Graphics (TOG)*, 37(6), pp.1-12.
- Jaderberg, M., Simonyan, K., Vedaldi, A. and Zisserman, A., 2014. Synthetic data and artificial neural networks for natural scene text recognition. *arXiv preprint arXiv:1406.2227*.
- Kniaz, V.V., Gorbatshevich, V.S. and Mizginov, V.A., 2017. Thermalnet: A deep convolutional network for synthetic thermal image generation. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 42, p.41.
- Li, Z., Zhang, L., Zhong, R., Fang, T., Zhang, L. and Zhang, Z., 2016. Classification of urban point clouds: A robust supervised approach with automatically generating training data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(3), pp.1207-1220.
- Malinverni, E.S., Pierdicca, R., Paolanti, M., Martini, M., Morbidoni, C., Matrone, F. and Lingua, A., 2019. Deep Learning for Semantic Segmentation of 3d Point Cloud. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*.
- Martínek, J., Lenc, L. and Král, P., 2019, May. Training Strategies for OCR Systems for Historical Documents. In *IFIP International Conference on Artificial Intelligence Applications and Innovations* (pp. 362-373). Springer, Cham.
- Miller Jr, E.F. and Melton, R.A., 1975, April. Automated generation of testcase datasets. In *Proceedings of the international conference on Reliable software* (pp. 51-58).
- Moreira, A., Quattrini, R., Maggiolo, G. and Mammoli, R., 2018. HBIM methodology as a bridge between Italy and Argentina. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, 42(2).
- Myers, G., 1999, August. A dataset generator for whole genome shotgun sequencing. In *ISMB* (pp. 202-210).
- Nespeca, R., 2019. Towards a 3D digital model for management and fruition of Ducal Palace at Urbino. An integrated survey with mobile mapping. *SCIRES-IT-SCientific REsearch and Information Technology*, 8(2), pp.1-14.
- Pei, Y. and Zaiane, O., 2006. A synthetic data generator for clustering and outlier analysis.
- Pierdicca, R., Marni, M., Malinverni, E.S., Paolanti, M. and Frontoni, E., 2019, June. Automatic Generation of Point Cloud Synthetic Dataset for Historical Building Representation. In *International Conference on Augmented Reality, Virtual Reality and Computer Graphics* (pp. 203-219). Springer, Cham.

Pierdicca, R., Paolanti, M., Matrone, F., Martini, M., Morbidoni, C., Malinverni, E.S., Frontoni, E. and Lingua, A.M., 2020. Point Cloud Semantic Segmentation Using a Deep Learning Framework for Cultural Heritage. *Remote Sensing*, 12(6), p.1005.

Qi, C.R., Su, H., Mo, K. and Guibas, L.J., 2017a. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 652-660).

Qi, C.R., Yi, L., Su, H. and Guibas, L.J., 2017b. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. In *Advances in neural information processing systems* (pp. 5099-5108).

Quattrini, R., Clini, P., Nespeca, R. and Ruggeri, L., 2016. Measurement and Historical Information Building: challenges and opportunities in the representation of semantically structured 3D content. *Disegnarecon*, 9(16), pp.14-1.

Quattrini, R., Pierdicca, R., Morbidoni, C. and Malinverni, E.S., 2017. Conservation-oriented HBIM. the BIMEXPLORER web tool. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, 42.

Ros, G., Sellart, L., Materzynska, J., Vazquez, D. and Lopez, A.M., 2016. The synthia dataset: A large collection of synthetic images for semantic segmentation of urban scenes. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3234-3243).

Tran, H. and Khoshelham, K., 2019. Building Change Detection Through Comparison of a Lidar Scan with a Building Information Model. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*.

Wang, F., Zhuang, Y., Gu, H. and Hu, H., 2019a. Automatic generation of synthetic LiDAR point clouds for 3-d data analysis. *IEEE Transactions on Instrumentation and Measurement*, 68(7), pp.2671-2673.

Wang, T., Wu, D.J., Coates, A. and Ng, A.Y., 2012, November. End-to-end text recognition with convolutional neural networks. In *Proceedings of the 21st international conference on pattern recognition (ICPR2012)* (pp. 3304-3308). IEEE.

Wang, Y., Sun, Y., Liu, Z., Sarma, S.E., Bronstein, M.M. and Solomon, J.M., 2019b. Dynamic graph cnn for learning on point clouds. *Acm Transactions on Graphics (tog)*, 38(5), pp.1-12.

Wu, B., Zhou, X., Zhao, S., Yue, X. and Keutzer, K., 2019, May. Squeezesegv2: Improved model structure and unsupervised domain adaptation for road-object segmentation from a lidar point cloud. In *2019 International Conference on Robotics and Automation (ICRA)* (pp. 4376-4382). IEEE.

Xu, Z., Sunkavalli, K., Hadap, S. and Ramamoorthi, R., 2018. Deep image-based relighting from optimal sparse samples. *ACM Transactions on Graphics (TOG)*, 37(4), pp.1-13.

6.2 Ringraziamenti e appendici opzionali

Questa ricerca è parzialmente finanziata dal progetto CIVITAS (Chain for excellence of reflective society to exploit digital cultural heritage and museums) (Clini et al., 2020). Gli autori

desiderano ringraziare il Prof. Clini per aver fornito i dati e i modelli del rilievo del Palazzo Ducale di Urbino e di Palazzo Ferretti, nonché i precedenti modelli 3D realizzati nella Palladio Library.

APPROCCIO GERARCHICO DI MACHINE LEARNING PER LA SEGMENTAZIONE SEMANTICA DI NUVOLE DI PUNTI 3D

A HIERARCHICAL MACHINE LEARNING APPROACH FOR POINT CLOUD SEGMENTATION

Eleonora Grilli^a, Simone Teruggi^b, Michele Russo^c, Francesco Fassi^b, Fabio Remondino^a

^a 3D Optical Metrology (3DOM) unit, Bruno Kessler Foundation (FBK), Via Sommarive 18, 38121 Trento, Italy; grilli@fbk.eu (E.G.), remondino@fbk.eu (F.R.)

^b 3DSurveyGroup – ABCLab – Politecnico di Milano, via Ponzio 31, 20133 Milano, Italy; simone.teruggi@polimi.it (S.T.), francesco.fassi@polimi.it (F.F.)

^c Department of History, Representation and Restoration of Architecture, Sapienza University of Rome, Rome, Italy Email: m.russo@uniroma1.it (M.R.)

PAROLE CHIAVE: nuvole di punti, classificazione, suddivisione gerarchica, machine learning

KEY WORDS: point cloud, classification, hierarchical segmentation, machine learning

RIASSUNTO

L'uso di dati 3D, nuvole di punti e mesh, per la documentazione, la valorizzazione e la visualizzazione del patrimonio è diventato sempre più diffuso. Ricchi di informazioni metriche, questi dati 3D soffrono la mancanza di informazioni strutturate quali la semantica e la gerarchia tra le parti. In questo contesto, l'introduzione di metodi automatici di classificazione può svolgere un ruolo essenziale per permettere un utilizzo reale di questi dati nelle operazioni di manutenzione e conservazione del bene culturale, agevolando un migliore utilizzo dei dati ai fini informativi e di analisi. In questo articolo viene presentato un innovativo approccio di classificazione multilivello e multi-risoluzione (MLMR). L'approccio MLMR proposto migliora il processo di apprendimento e ottimizza i risultati della classificazione 3D attraverso un concetto gerarchico. La procedura MLMR viene testata e valutata su due diversi datasets, complessi e di grandi dimensioni: l'Abbazia di Pomposa (Italia) e il Duomo di Milano (Italia). I risultati della classificazione mostrano l'affidabilità e la replicabilità del metodo sviluppato, permettendo l'identificazione di svariate classi architettoniche a diversi livelli di risoluzione geometrica.

ABSTRACT

The recent years saw an extensive use of 3D point cloud data for heritage documentation, valorisation, and visualisation. Although rich in metric quality, these 3D data lack structured information such as semantics and hierarchy between parts. In this context, the introduction of point cloud classification methods can play an essential role for better data usage, model definition, analysis, and conservation. The paper aims to extend a machine learning (ML) classification method with a multi-level and multi-resolution (MLMR) approach. The proposed MLMR approach improves the learning process and optimises 3D classification results through a hierarchical concept. The MLMR procedure is tested and evaluated on two large-scale and complex datasets: the Pomposa Abbey (Italy) and the Milan Cathedral (Italy). Classification results show the reliability and replicability of the developed method, allowing the identification of the necessary architectural classes at each geometric resolution.

1. INTRODUZIONE

La segmentazione semantica/classificazione di nuvole di punti è un passaggio cruciale nel trattamento dei dati rilevati. Essa può assumere un ruolo importante nel processo di definizione ed analisi dei dati 3D in forma discreta, usandoli nella pratica quotidiana manutentiva dei Beni Culturali e, più in generale, in tutte quelle operazioni legate alla conoscenza e al restauro della architettura. Il termine segmentazione semantica fa riferimento al raggruppamento in sottoinsiemi di dati simili (detti segmenti), i quali presentano caratteristiche/features, (geometriche, radiometriche, ecc.), attraverso le quali è possibile distinguere e identificare le diverse parti che compongono la nuvola di punti. La suddivisione semantica del dato 3D proveniente dal rilievo, può portare ad una gerarchizzazione del dato geometrico e radiometrico, che permette un uso diretto della nuvola di punti nel processo decisionale e di progetto (ad oggi molto limitato) ed è propedeutica alla fase di ricostruzione di modelli 3D

semplificati (CAD o BIM) e a alle classiche restituzioni 2D. Ciò a sua volta comporta una semplificazione nella lettura e una più semplice gestione dei dati acquisiti, accelerando la fase di interpretazione e di ricostruzione del dato vettoriale. Partendo direttamente dalla nuvola di punti classificata è possibile inoltre avviare un percorso di analisi finalizzato ad un incremento della conoscenza della architettura propedeutico alla progettazione di interventi manutentivi o conservativi, senza bisogno di ulteriori elaborazioni spesso dispendiose e complesse. In tal senso, la possibilità di individuare ed evidenziare a priori le principali componenti architettoniche consente una più facile lettura e comprensione da parte di un pubblico più ampio, incrementando la possibilità di interazione e il coinvolgimento in progetti che riguardano il costruito. Ma la capacità di poter separare, visualizzare e gestire separatamente elementi morfologicamente omogenei permette anche una più rapida analisi spaziale del manufatto, potendo effettuare studi sui singoli elementi, verificarne il dimensionamento e la collocazione nello spazio

secondo regole modulari, analizzando e confrontando gli elementi ricorrenti e le singolarità. Dal punto di vista dell'analisi statica del manufatto, la possibilità di estrarre e classificare separatamente gli elementi della nuvola consente di effettuare uno studio visivo fondato sull'astrazione degli elementi strutturali, preliminare ad una conseguente simulazione con sistemi FEM/FEA. Infine, la possibilità di tradurre questo sistema ordinato di informazioni geometriche in un abaco di modelli poligonali 3D offre l'opportunità di costruire un sistema al quale poter associare informazioni di carattere morfologico e materico, estremamente utile sia in caso di pianificazione conservativa che manutentiva dell'architettura. La possibilità di definire aree, volumi e masse, nel caso di materiali omogenei, permette di quantificare l'intero manufatto, producendo un computo metrico del bene culturale, ampliando la conoscenza di ogni singolo dettaglio e fornendo uno strumento molto utile per il progetto di restauro. Da questo risulta evidente l'utilità nel poter raggiungere in breve tempo la segmentazione semantica del dato 3D rilevato, soprattutto se si tratta di architetture articolate, poiché tale attività semplifica il complesso processo di conoscenza e manutenzione del bene culturale.

1.1 Scopo della ricerca

In questo articolo viene presentato e valutato un metodo di classificazione automatica, basato su un approccio multi-livello e multi-risoluzione (MLMR) combinato con un algoritmo di apprendimento automatico (Machine Learning). Il metodo MLMR classifica gerarchicamente i dati 3D a diverse risoluzioni geometriche per facilitare il processo di apprendimento e ottimizzare i risultati della classificazione. Per testare la metodologia sviluppata sono stati presi in considerazione due dataset complessi e di grande estensione, quali il Duomo di Milano e l'Abbazia di Pomposa. Entrambi i casi studio presentano grande diversità di geometrie e stili, ricchezza di decorazioni e mancanza di regolarità degli elementi architettonici. La scelta di due strutture architettoniche così diverse e complesse è volta a dimostrare l'applicabilità della metodologia a diversi scenari.

2. STATO DELL'ARTE

Nel campo del patrimonio architettonico, la necessità di strutturare e classificare i dati di rilievo 3D è stata sperimentata in diverse applicazioni come strumento di supporto dei processi scan-to-BIM (Rebolj et al., 2017; Son and Kim, 2017; Bassier and Yousefzadeh, 2020), di attività di monitoraggio e di restauro (Apollonio et al., 2018; Croce et al., 2020), pianificazione di opere di manutenzione (Sanchez et al., 2018; Valero et al., 2018; Roussel et al., 2019), o ancora rilevamento dei danni (Mizoguchi et al., 2013; Kashani and Graettinger, 2015).

Tuttavia, nella maggior parte dei casi, la suddivisione manuale dei dataset, oltre ad essere un intervento laborioso e dispendioso, porta con sé un grado elevato di soggettività e di non ripetibilità e di non riproducibilità del processo. Negli ultimi anni sono stati compiuti progressi rilevanti nei processi di classificazione automatica mediante l'utilizzo di metodi di Intelligenza Artificiale, Machine Learning (ML) e Deep Learning (DL), che, al contrario, sono oggettivi, replicabili e ripetibili. Le tecniche standard di ML supervisionato prevedono che gli algoritmi prendano come input alcune parti manualmente annotate della nuvola di punti, assieme alle cosiddette "features", attributi geometrici e/o radiometrici selezionati dall'operatore per facilitare l'apprendimento e la distinzione delle classi ricercate. Dall'altro lato, le strategie di DL prevedono la generazione automatica delle features, che apprendono grazie all'utilizzo di

grandi quantità di dati annotati in input. Per far fronte alla mancanza di dati architettonici utili per l'addestramento degli algoritmi è stata recentemente rilasciata una raccolta (ArCH) di nuvole punti architettoniche annotate (Matrone et al., 2020a). Per facilitare il processo di annotazione, Murtiyoso e Grussenmeyer (2020) hanno sviluppato una serie di funzioni matematiche tramite le quali è possibile isolare alcune specifiche classi architettoniche all'interno della nuvola di punti, come ad esempio colonne e travi. Nell'ambito dei beni culturali, Grilli et al. (2018) hanno proposto un approccio Machine Learning di tipo supervisionato che trasferisce le informazioni di classificazione dalle texture 2D ai modelli 3D. Gli stessi autori hanno poi applicato l'addestramento supervisionato direttamente sulle nuvole di punti formando un classificatore Random Forest (RF) con features geometriche studiate ad hoc per l'architettura (Grilli et al., 2019; Grilli e Remondino, 2020).

In ambito DL, Pierdicca et al., (2020) hanno "addestrato" sul dataset ArCH un Dynamic Graph Convolutional Neural Network (DGCNN) supportato da features significative (colori, normali e HSV). Il modello risultante è stato poi testato in due modi diversi: su un dataset parzialmente etichettato e su una scena non vista, fornendo risultati promettenti.

Recentemente è stato inoltre pubblicato un articolo di confronto sulla classificazione di nuvole di punti architettoniche utilizzando tecniche ML o DL (Matrone et al., 2020b). Dall'articolo è emerso che i risultati ottenuti con ML e DL sono comparabili in termini di accuratezza, sebbene i tempi impiegati nel ML siano ancora di gran lunga inferiori rispetto al DL.

Considerata la complessità e peculiarità dei casi studio trattati in questo articolo, si è scelto di prendere come punto di partenza l'approccio di ML proposto da (Grilli e Remondino, 2020), in quanto risulta facilmente adattabile a nuove classi architettoniche durante la fase di addestramento dell'algoritmo.

3. CASI STUDIO

3.1 Duomo di Milano

A supporto delle attività di manutenzione straordinaria della cattedrale tardo-gotica, negli ultimi dieci anni sono state condotte diverse operazioni di rilievo della Guglia Maggiore, degli altari del transetto e del tiburio, che hanno portato alla produzione di modelli 3D di tipo nurbs e rappresentazioni 2D in scala 1:50 (Fassi et al., 2011). Negli ultimi tre anni inoltre, combinando l'uso di scansioni laser terrestri per gli spazi interni e di fotogrammetria per quelli esterni, è stata generata una nuvola di punti 3D dell'intera Cattedrale ad una risoluzione media uniforme di 5 mm (Achille et al., 2020).

Dati i necessari lavori di manutenzione continua e la grande quantità di dati da gestire, la segmentazione semantica della nuvola di punti del Duomo di Milano può diventare un utile strumento di supporto digitale delle attività di conservazione evitando lunghi processi di modellazione che diventano costosi, complessi, spesso non sostenibili ed inoltre degradano il dato 3D in termini di accuratezza, affidabilità metrica e di risoluzione. L'utilizzo diretto nella pratica decisionale e di progetto della nuvola di punti permetterebbe un'immediata derivazione di misure e rappresentazioni 2D; l'identificazione, conteggio e visualizzazione dei singoli elementi architettonici per una migliore interpretazione delle strutture architettoniche a livello di nuvola di punti. Favorirebbe inoltre la possibilità di tracciamento di ogni attività di restauro, trattando la nuvola di punti come un completo sistema informativo navigabile in 3D dove è possibile consultare informazioni, dati e un catalogo di documenti d'archivio generando una piattaforma informativa web-based di

tipo BIM, utilizzabile direttamente sul campo o all'interno di un sistema a realtà mista.

3.2 Abbazia di Pomposa

Il progetto di rilievo del monastero benedettino dell'Abbazia di Pomposa, risalente al IX secolo, è stato eseguito nel 2015 mediante l'utilizzo di tecnologie di tipo attivo e passivo, terrestri ed aeree, al fine di valutare la possibilità di integrare tecniche di acquisizione fotogrammetrica basate sul processo di Structure from Motion (SfM) con consolidate metodologie di acquisizione 3D da laser scanner, ottenendo in tal senso un modello integrato completo dell'intero complesso abbaziale. La campagna di acquisizione 3D del complesso monastico ha replicato una attività conoscitiva digitale già avviata nel 1998, attraverso un progetto pionieristico per le metodologie di rilevamento adottate. La campagna attuale si è fondata sulla interazione di strumenti attivi e passivi, da terra e da drone (Russo e Manfredini, 2014). Nel caso dell'abbazia l'arricchimento della nuvola di punti con informazioni semantiche può essere utile per differenti finalità. In primis è possibile derivare velocemente da queste informazioni delle misure e relative rappresentazioni 2D. Inoltre, è possibile pianificare un monitoraggio dell'edificio nel tempo, verificando murature, colonne e rotazioni delle coperture: un confronto tra nuvole di punti acquisite nel tempo va considerato come una valida soluzione per l'esatta e completa descrizione dell'intero tessuto. Per quanto riguarda il controllo geometrico sulla qualità dei dati acquisiti, questo può avvenire sui singoli elementi appartenenti alla stessa classe architettonica, evidenziando possibili alterazioni della composizione strutturale dell'edificio. Infine, tale suddivisione può facilitare l'operazione di "quantificazione" dell'edificio dal punto di vista materico e funzionale, raccogliendo alcune informazioni molto importanti sia per la pianificazione di un intervento conservativo che per una eventuale valutazione dei danni (ad esempio dopo un evento distruttivo).

4. METODOLOGIA

La complessità dei casi studio scelti ci pone principalmente davanti a due problemi:

- la dimensione del dato rende il processo di calcolo impegnativo in termini computazionali. Un sottocampionamento della nuvola di punti potrebbe aiutare a gestire l'elevato numero di punti, ma abbasserebbe il livello di dettaglio del dataset originale;
- il numero elevato di classi semantiche da identificare al contempo induce rapidamente ad errori di classificazione: alcuni esperimenti iniziali hanno dimostrato che più alto è il numero di classi, minore è l'accuratezza della classificazione (Teruggi et al., 2020).

Considerando i problemi sopra citati, si è quindi ritenuto opportuno sviluppare la metodologia presentata da Grilli et al. (2020), in cui i dati venivano classificati in un unico passaggio, suddividendo ora il processo su più fasi. Si propone quindi un approccio multi-livello e multi-risoluzione (MLMR), che segue i seguenti passaggi:

1. La nuvola di punti a risoluzione completa viene sottocampionata e classificata tramite apprendimento supervisionato nelle sue parti principali (macro-categorie).
2. I risultati provenienti da questa classificazione vengono interpolati su una versione a più alta risoluzione della nuvola di punti, trasferendo le classi assegnate in prima fase in un determinato intorno di punti.

3. le macrocategorie di interesse vengono a sua volta suddivise nelle loro componenti (ad esempio, le colonne in base, fusto e capitello);
4. i risultati della classificazione vengono nuovamente interpolati su una versione a più alta risoluzione della nuvola di punti fino a raggiungere la piena risoluzione geometrica.

Va detto che ogni elemento può essere classificato in modo sempre più dettagliato, in base alla risoluzione della nuvola di punti usata al livello corrispondente, e poi interpolato su una versione a risoluzione più alta. Il processo è iterativo, e l'ultimo livello di classificazione corrisponderà alla nuvola di punti a piena risoluzione. Inoltre, il sotto campionamento della nuvola di punti dipenderà sia dalla complessità dell'oggetto da classificare che dagli elementi da riconoscere. Le macro-categorie di oggetti, i macro-elementi architettonici, consentono una risoluzione più bassa, mentre la risoluzione più alta viene utilizzata per gli elementi di dettaglio.

4.1 Selezione delle features

Le features geometriche selezionate per addestrare il modello di ML al riconoscimento automatico degli elementi architettonici sono le cosiddette "features di covarianza" (Blomey et al., 2014), le quali consentono di evidenziare le caratteristiche geometriche locali del dato 3D. Vengono, in questo modo, evidenziati i comportamenti locali della nuvola come ad esempio la "linearità", "planarità" e "sfericità" o ancora la "variazione di superficie" e l'"anisotropia". La scelta delle giuste features geometriche ed il raggio di intorno del punto sul quale calcolarle sono fondamentali ai fini dell'individuazione degli elementi architettonici (Figura 1) (Grilli et al., 2019).

La metodologia gerarchica proposta è sintetizzata graficamente nella Figura 5. Il diagramma ha lo scopo di fornire un'indicazione generale in termini di risoluzione geometrica della nuvola di punti e di features geometriche da utilizzare in ogni fase della classificazione. I parametri (risoluzione e raggi delle features), possono variare caso per caso, a seconda delle caratteristiche e delle dimensioni degli oggetti. Tuttavia, si può notare che essi corrispondono al più piccolo dettaglio che può essere rappresentato ad una data scala di rappresentazione e alla sua tolleranza metrica, in una logica di rappresentazione ad un livello di dettaglio crescente.

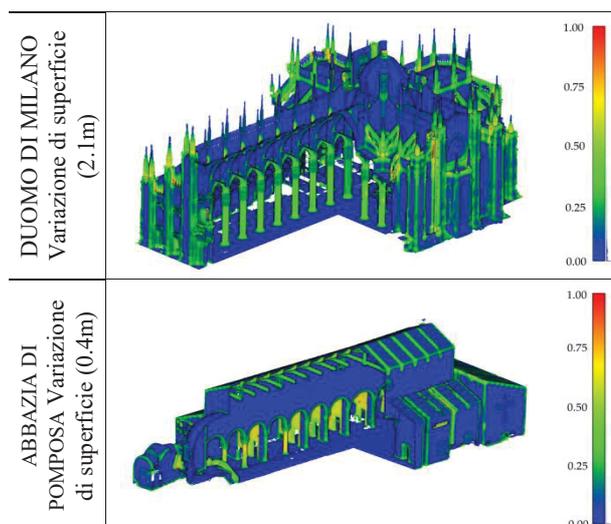


Figura 1. Utilizzo della stessa feature geometrica, calcolata a raggi diversi poiché selezionata in funzione delle dimensioni del caso studio trattato.

4.2 Annotazione del dato

In aggiunta alla selezione delle features geometriche, alcune parti di ogni dataset, ad ogni livello di classificazione, devono essere annotate manualmente per addestrare l'algoritmo Random Forest (Breiman, 2001). La Figura 2 mostra i set di addestramento utilizzati al primo livello di classificazione (risoluzione 5 cm) del Duomo di Milano e dell'Abbazia di Pomposa, rispettivamente. Per il primo caso studio (30 milioni di punti), il campione di addestramento è composto da circa 2,5 milioni di "punti etichettati" per le superfici esterne e circa 2,6 milioni di punti per gli spazi interni. Per l'Abbazia di Pomposa, di 1,1 milione di punti che compongono l'intera nuvola circa 115.444 punti sono stati classificati manualmente.

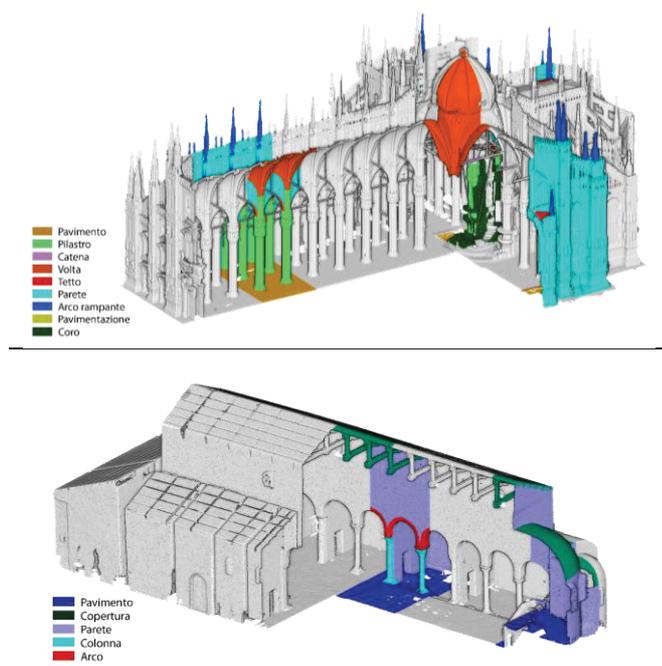


Figura 2. Porzioni di dataset annotate manualmente ai fini di addestramento al livello 1.

4.3 Valutazione dei risultati

Per valutare quantitativamente le prestazioni della classificazione, sono state prese in considerazione le metriche di accuratezza comunemente utilizzate in ambito machine learning, quali "Precision", "Recall" e "F1 score" (Goutte e Gaussier, 2005). Tali metriche vengono ricavate dal confronto diretto tra predizione automatica e annotazione manuale della stessa porzione di nuvola e tengono conto non solo dei valori correttamente predetti ma anche di quelli erronei.

5. RISULTATI

La tabella 1 riassume i tempi di addestramento e di classificazione impiegati per i due casi studi ai tre livelli di classificazione. In aggiunta la Figura 3 indica la percentuale di tempo trascorso in ogni fase di lavoro. È importante sottolineare che il tempo di calcolo per l'estrazione delle features geometriche varia in relazione alla dimensione della nuvola di punti, al numero di features e al loro raggio di ricerca. Per quanto riguarda il primo livello di classificazione, su una workstation con processore a 18 core, la fase di addestramento per il Duomo di Milano ha richiesto circa 5 minuti (2,5 mil. di punti), mentre sono stati necessari 43 secondi per classificare la restante nuvola di

punti (12 mil. di punti). Per quanto riguarda l'Abbazia di Pomposa, il processo di addestramento e di classificazione hanno richiesto rispettivamente 5 e 3 secondi.

		TEMPI (sec)		
		L1	L2	L3
DUOMO	Addestramento	363	17	142
	Classificazione	43	12	174
ABBZIA	Addestramento	5	1	4
	Classificazione	2.7	1	29

Tabella 1. Riepilogo dei tempi impiegati per addestramento e classificazione nei tre livelli di classificazione.

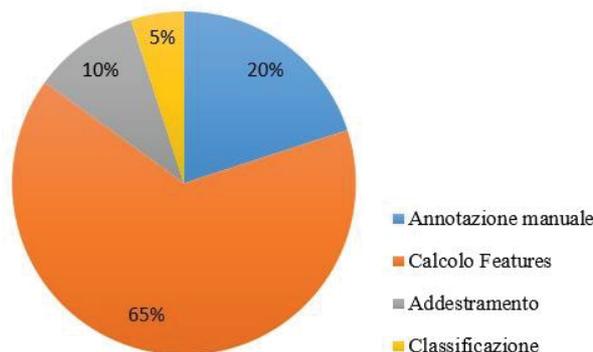


Figura 3. Rappresentazione sintetica dei tempi (normalizzati) necessari per le diverse fasi del processo di classificazione, dall'annotazione manuale alla classificazione finale del dataset.

5.1 Duomo di Milano

Applicando l'approccio gerarchico, la nuvola di punti del Duomo è stata classificata su tre livelli (Fig. 6 -7). Al primo livello la classificazione è stata eseguita su nuvola di punti con risoluzione di 5 cm, calcolando features geometriche di raggio compreso tra 20 cm e 2,5 m, permettendo l'identificazione di macro-elementi architettonici. Al secondo livello di classificazione, dopo aver trasferito la classificazione proveniente dal primo livello su nuvola di punti con risoluzione di 2 cm, gli elementi architettonici precedentemente identificati sono stati suddivisi in sottocomponenti, riducendo a 10 cm e 1 m i raggi min./max di ricerca delle features geometriche.

Infine, il terzo livello di classificazione mira alla suddivisione di ogni componente nei suoi conci di marmo (elementi monolitici quali statue, decorazioni gotiche, ecc.). In questo caso, quindi, è stata utilizzata la nuvola di punti a piena risoluzione, ricercando le features con raggio minimo e massimo di 0,5 e 5 cm rispettivamente. La Tabella 2 riporta i risultati in termine di F1 score ottenuti nei tre livelli, con una precisione sempre superiore al 90%.

	Level 1 (5cm)	Level 2 (2 cm)	Level 3 (0.5 cm)
Precision (%)	94.7	99	92
Recall (%)	95	98	88.5
F1 score (%)	93.78	99.3	91.8

Tabella 2. Metriche di accuratezza ottenute ai tre livelli di classificazione del Duomo di Milano.

6. CONCLUSIONI

In questo articolo è stata presentata una nuova strategia di classificazione gerarchica (MLMR) che permette di assegnare un significato semantico alle nuvole di punti 3D di strutture complesse appartenenti al mondo dei beni culturali, processo classicamente manuale in fase di modellazione 3D.

Gli aspetti innovativi del lavoro presentato sono i seguenti:

- L'uso di tecniche di machine learning per classificare rapidamente grandi e complesse architetture 3D senza la necessità di annotare grandi quantità di dati di training.
- La definizione di regole generali replicabili in vari scenari del patrimonio, in termini di relazioni tra i livelli di classificazione, risoluzione delle nuvole di punti e raggi di ricerca delle features minime/massime.
- La segmentazione gerarchica (fino a singole istanze) dei dati 3D, strumento utile per la modellazione HBIM.
- La velocità del processo: una volta definiti i dataset di formazione e validazione, la previsione dell'intero set di dati si ottiene in pochi minuti.

I risultati di classificazione ottenuti mostrano come le nuvole di punti classificate possano facilitare la comprensione di complesse strutture architettoniche, diventando possibili strumenti di supporto al restauro, la manutenzione e la gestione delle strutture in loco senza il bisogno di sintetizzazioni successive. Sebbene il processo sia automatico il contributo cognitivo di un operatore esperto è fondamentale all'inizio del processo. Proprio perché i Beni Culturali sono variegati non si possono fare generalizzazioni assolute, all'operatore esperto che conosce il Bene in esame, è richiesto lo sforzo della definizione delle classi e la scelta dei set di formazione e di convalida (annotazione dei dati). Queste fasi sono cruciali per adattare il processo ai diversi casi di studio e alle diverse finalità.

Come possibili linee di ricerca future, alcuni aspetti potrebbero meritare ulteriore attenzione e sviluppo:

- una migliore indagine sulla relazione tra i livelli di classificazione, la risoluzione delle nuvole di punti e i raggi di ricerca delle features: è necessario capire se la classificazione automatica con features specifiche può essere generalizzata per quanto riguarda la densità dei dati, o se dipende dal caso;
- verifica dell'utilità del processo di classificazione per il processo scan-to-BIM, verificando se le strutture e le istanze semantiche estratte facilitano il lavoro preparatorio per la costruzione di modelli BIM;
- verifica se le nuvole di punti semanticamente segmentate possono facilitare la generazione di mesh poligonali;
- Creazione di un quadro di classificazione di più facile utilizzo per i non esperti del settore;
- Verifica della possibilità di elaborare automaticamente i dati acquisiti in loco con strumenti scanner mobili per applicazioni di monitoraggio in tempo reale.
- Miglioramento dei dettagli di classificazione integrando le informazioni provenienti da immagini, che in genere hanno una risoluzione più elevata, permettendo così una migliore identificazione/distinzione dei piccoli elementi (es. classificazione di ogni singolo blocco di marmo che compone il Duomo di Milano).

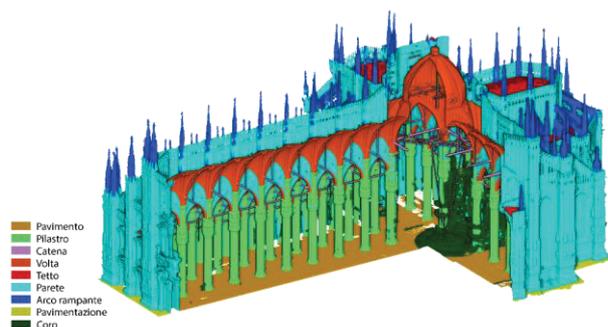


Figura 3. Risultato ottenuto dalla classificazione automatica del Duomo al Livello 1.

5.2 Abbazia di Pomposa

Considerando le esigenze di monitoraggio, la classificazione è stata suddivisa in tre diversi livelli (Figura 8 - 9). Il primo livello fa riferimento al quadro strutturale principale della chiesa, definendo una suddivisione funzionale globale dell'intero dataset: pavimento, facciate, colonne, archi e tetto. Per riconoscere queste macro-categorie, la nuvola di punti è stata elaborata con una risoluzione di 5 cm.

Il secondo livello è dedicato ad un processo di classificazione più approfondito, riconoscendo più sottoclassi all'interno delle categorie provenienti dal primo livello di classificazione. Le facciate sono state suddivise in pareti e finestre, le colonne in basi, fusti e capitelli ed il tetto in cupole, copertura laterale, centrale e frontale. A questo livello, ogni elemento architettonico è stato riconosciuto sulla nuvola di punti con risoluzione di 2 cm. Il terzo livello è stato unicamente focalizzato allo studio delle coperture lignee, le quali presentano strutture complesse e variabili. Per ogni diversa struttura del tetto è stato eseguito un processo di classificazione diversificato, evidenziandone i principali elementi strutturali (ad esempio, travi, travetti, saette, puntoni, tavole, ecc.). Per completare questo studio strutturale, si è deciso di sfruttare nuovamente la nuvola con risoluzione 2 cm. La tabella 3 riporta le metriche di accuratezza ottenute.

	Level 1 (5cm)	Level 2 (2 cm)	Level 3 (0.5 cm)
Precision (%)	95.3	98	95.8
Recall (%)	95.1	97.7	95.7
F1 score (%)	95.1	97.8	94.6

Tabella 3. Metriche di accuratezza ottenute ai tre livelli di classificazione dell'Abbazia di Pomposa.

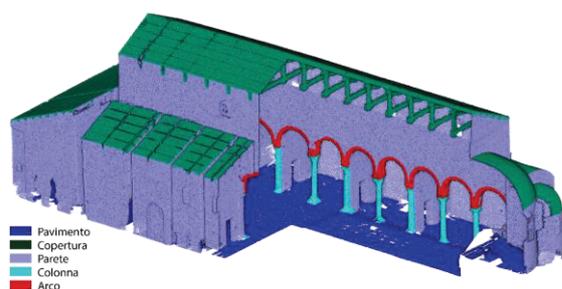


Figura 4. Risultato ottenuto dalla classificazione automatica dell'Abbazia al Livello 1.

BIBLIOGRAFIA

- Achille, C., Fassi, F., Mandelli, A., Perfetti, L., Rechichi, F. and Teruggi, S., 2020. From a Traditional to a Digital Site: 2008–2019. The History of Milan Cathedral Surveys. In *Digital Transformation of the Design, Construction and Management Processes of the Built Environment* (pp. 331-341). Springer, Cham.
- Apollonio, F.I., Basilissi, V., Callieri, M., Dellepiane, M., Gaiani, M., Ponchio, F., Rizzo, F., Rubino, A.R. and Scopigno, R., 2018. A 3D-centered information system for the documentation of a complex restoration intervention. *Journal of Cultural Heritage*, 29, pp.89-99.
- Bassier, M., Yousefzadeh, M. and Vergauwen, M., 2020. Comparison of 2D and 3D wall reconstruction algorithms from point cloud data for as-built BIM. *Journal of Information Technology in Construction (ITcon)*, 25(11), pp.173-192.
- Blomley, R., Weinmann, M., Leitloff, J. and Jutzi, B., 2014. Shape distribution features for point cloud analysis—a geometric histogram approach on multiple scales. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2(3), 9.
- Breiman, L., 2001. Random forests. *Machine learning*, 45(1), pp. 5-32.
- Croce, V., Caroti, G., De Luca, L., Piemonte, A. and Véron, P., 2020. semantic annotations on heritage models: 2D/3D approaches and future research challenges. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 43, pp.829-836.
- Fassi, F., Achille, C. and Fregonese, L., 2011. Surveying and modelling the main spire of Milan Cathedral using multiple data sources. *The Photogrammetric Record*, 26(136), pp.462-487.
- Goutte, C. and Gaussier, E., 2005, March. A probabilistic interpretation of precision recall and F-score, with implication for evaluation. In *European conference on information retrieval* (pp. 345-359). Springer, Berlin, Heidelberg.
- Grilli, E. and Remondino, F., 2020. Machine Learning Generalisation across Different 3D Architectural Heritage. *ISPRS International Journal of Geo-Information*, 9(6), 379.
- Grilli, E., Dinunno, D., Marsicano, L., Petrucci, G. and Remondino, F., 2018. Supervised segmentation of 3D cultural heritage. In *2018 3rd Digital Heritage International Congress (DigitalHERITAGE) held jointly with 2018 24th International Conference on Virtual Systems & Multimedia (VSMM 2018)*, pp. 1-8. IEEE.
- Grilli, E., M. Farella, E., Torresani, A. and Remondino, F., 2019. Geometric features analysis for the classification of cultural heritage point clouds. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XLII-2/W15, pp 541-548.
- Kashani, A.G. and Graettinger, A.J., 2015. Cluster-based roof covering damage detection in ground-based lidar data. *Automation in Construction*, 58, pp.19-27.
- Matrone, F., Lingua, A., Pierdicca, R., Malinverni, E.S., Paolanti, M., Grilli, E., Remondino, F., Murtiyoso, A. and Landes, T., 2020a. A benchmark for large-scale heritage point cloud semantic segmentation. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XLIII-B2, pp. 1419-1426.
- Matrone, F., Grilli, E., Martini, M., Paolanti, M., Pierdicca, R. and Remondino, F., 2020b. Comparing Machine and Deep Learning Methods for Large 3D Heritage Semantic Segmentation. *ISPRS International Journal of Geo-Information*, 9(9), 535.
- Mizoguchi, T., Koda, Y., Iwaki, I., Wakabayashi, H., Kobayashi, Y., Shirai, K., Hara, Y. and Lee, H.S., 2013. Quantitative scaling evaluation of concrete structures based on terrestrial laser scanning. *Automation in construction*, 35, pp.263-274.
- Murtiyoso, A. and Grussenmeyer, P., 2020. Virtual Disassembling of Historical Edifices: Experiments and Assessments of an Automatic Approach for Classifying Multi-Scalar Point Clouds into Architectural Elements. *Sensors*, 20(8), 2161.
- Pierdicca, R., Paolanti, M., Matrone, F., Martini, M., Morbidoni, C., Malinverni, E.S., Frontoni, E. and Lingua, A.M., 2020. Point Cloud Semantic Segmentation Using a Deep Learning Framework for Cultural Heritage. *Remote Sensing*, 12(6), 1005.
- Rebolj, D., Pučko, Z., Babič, N.Č., Bizjak, M. and Mongus, D., 2017. Point cloud quality requirements for Scan-vs-BIM based automated construction progress monitoring. *Automation in Construction*, 84, pp.323-334.
- Roussel R., Bagnérís M., De Luca L. and Bomblet P., 2019. A digital diagnosis for the <<autumn>> statue (Marseille, France): photogrammetry, digital cartography and construction of a thesaurus. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XLII-2/W15, pp. 1039–1046.
- Russo, M. and Manferdini, A.M., 2014. Integration of image and range-based techniques for surveying complex architectures. *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences*, 2(5).
- Sánchez-Aparicio, L.J., Del Pozo, S., Ramos, L.F., Arce, A. and Fernandes, F.M., 2018. Heritage site preservation with combined radiometric and geometric analysis of TLS data. *Automation in Construction*, 85, pp.24-39.
- Son, H. and Kim, C., 2017. Semantic as-built 3D modeling of structural elements of buildings based on local concavity and convexity. *Advanced Engineering Informatics*, 34, pp.114-124.
- Teruggi, S., Grilli, E., Russo, M., Fassi, F. and Remondino, F., 2020. A Hierarchical Machine Learning Approach for Multi-Level and Multi-Resolution 3D Point Cloud Classification. *Remote Sensing*, 12(16), 2598.
- Valero, E., Bosché, F. and Forster, A., 2018. Automatic segmentation of 3D point clouds of rubble masonry walls, and its application to building surveying, repair and maintenance. *Automation in Construction*, 96, pp.29-39.

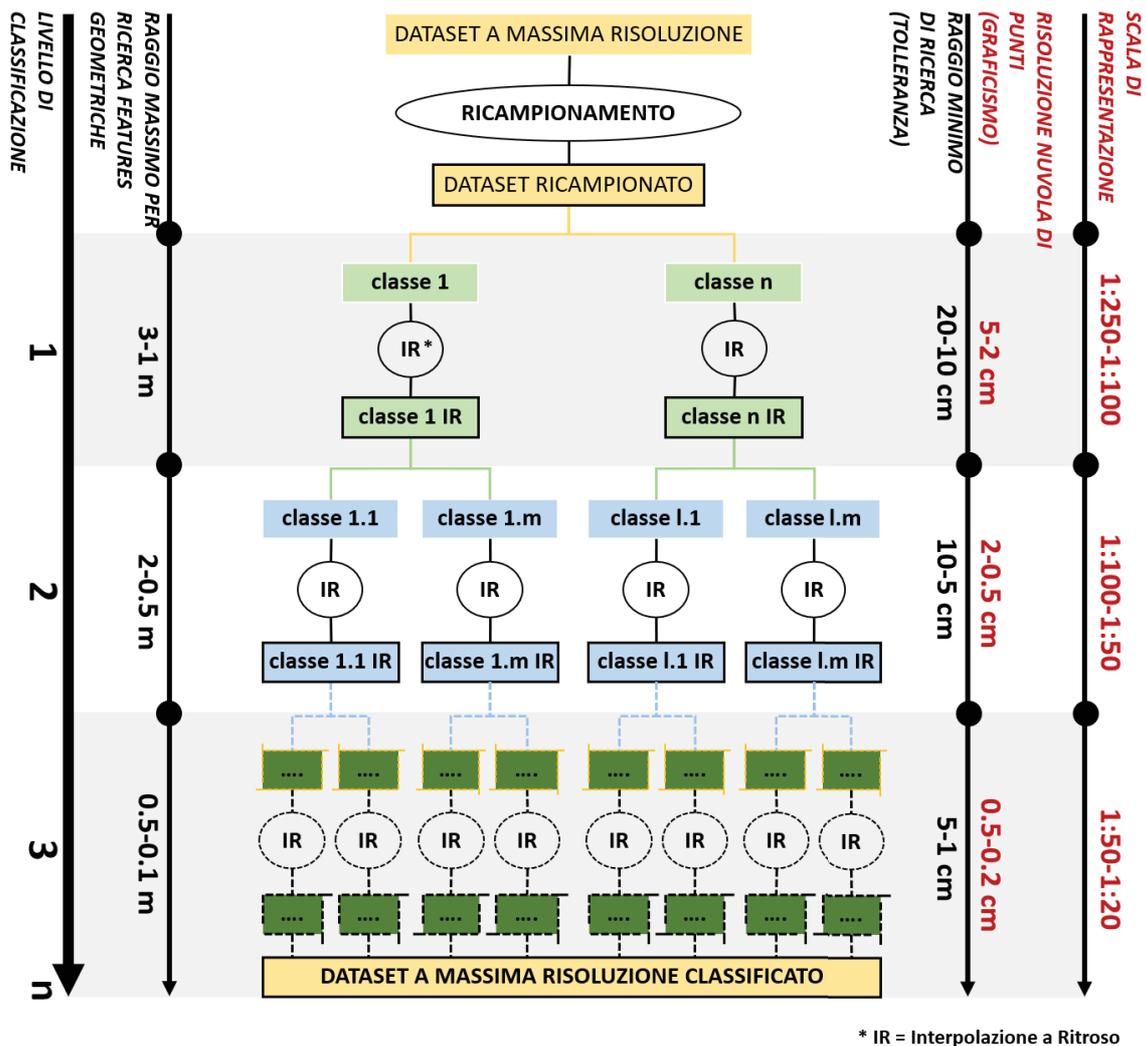
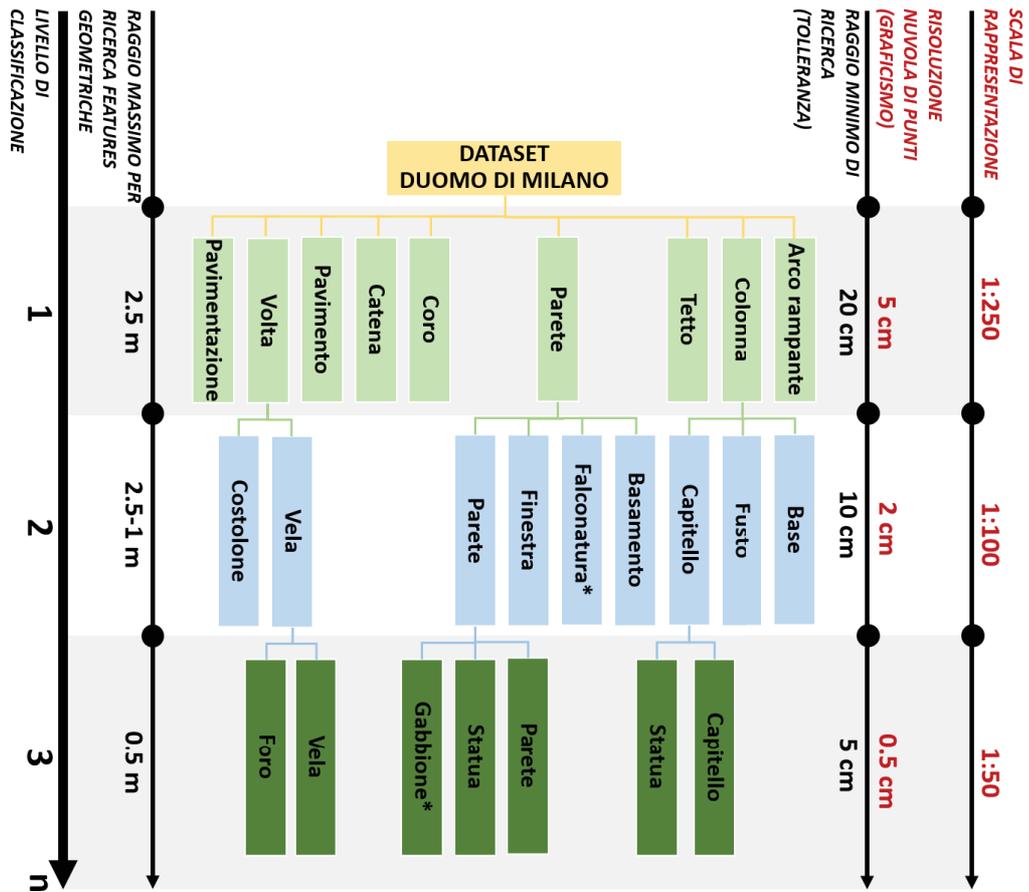


Figura 5. Approccio di classificazione a più livelli e risoluzioni (MLMR). Il diagramma fornisce indicazioni generali in termini di risoluzione delle nuvole di punti e raggio di ricerca minimo/massimo delle features geometriche, che devono essere scelte in ogni fase del processo di classificazione.



*Elemento decorativo della facciata.

Figura 6. Approccio di classificazione MLMR. Distribuzione delle classi ai diversi livelli di risoluzione per il Duomo di Milano.

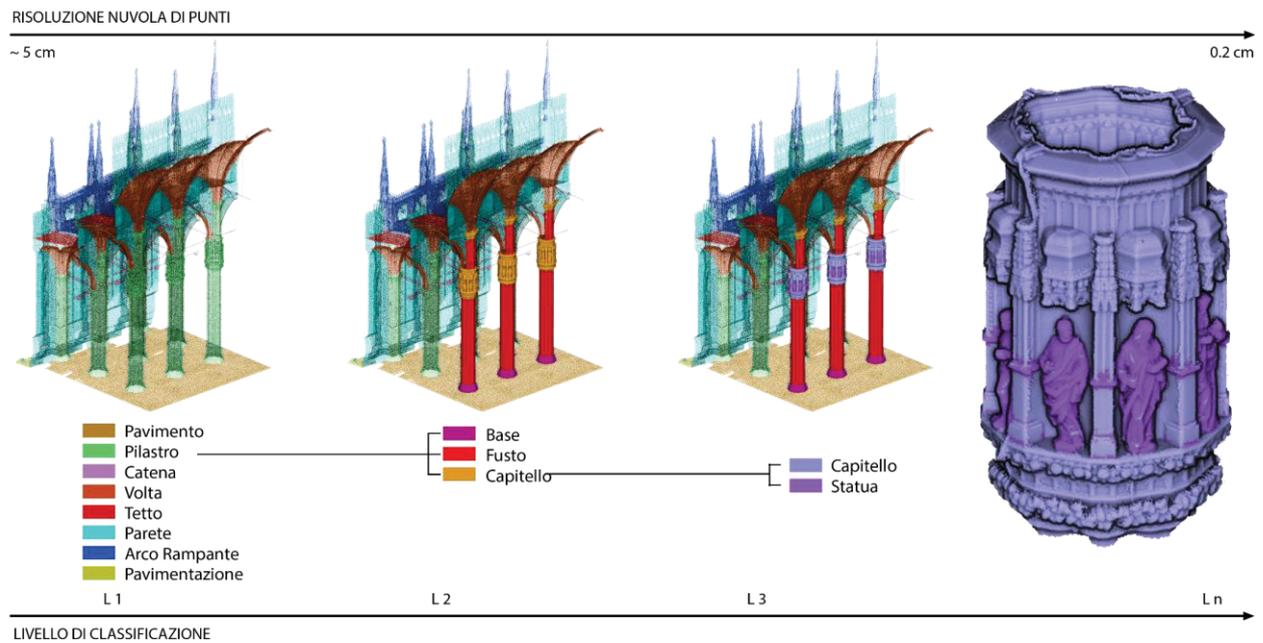


Figura 7. Approccio MLMR di classificazione. Risultati ottenuti sul Duomo di Milano.

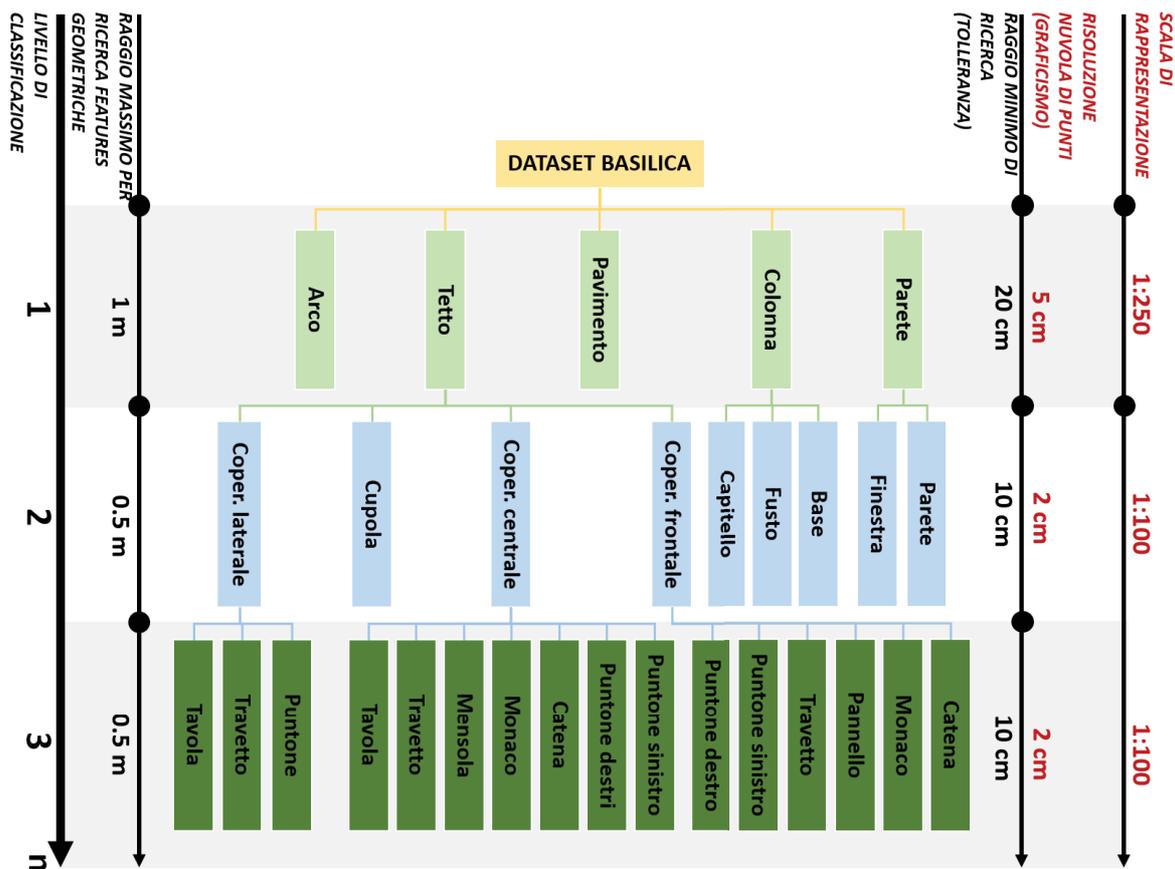


Figura 8. Approccio di classificazione MLMR. Distribuzione delle classi ai diversi livelli di risoluzione per l'Abbazia di Pomposa.

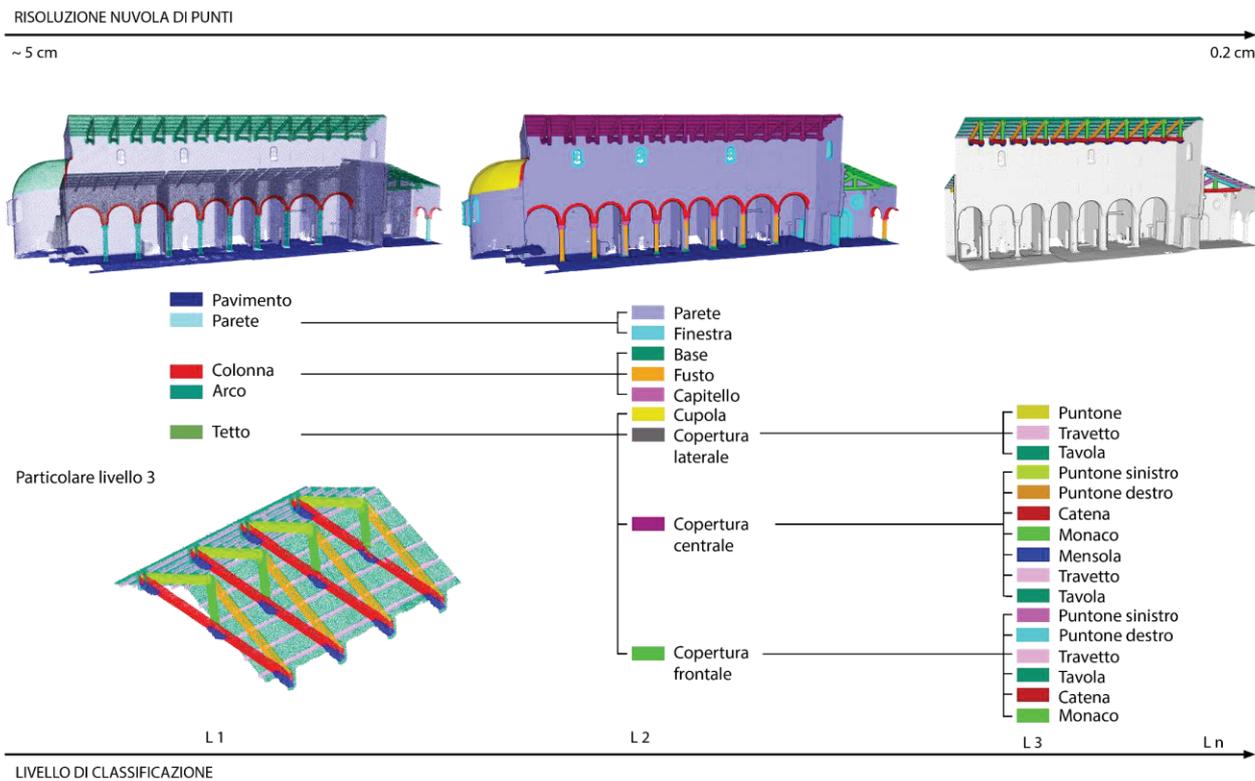


Figura 9. Approccio MLMR di classificazione. Risultati ottenuti sull'Abbazia di Pomposa.

ONTOLOGIE GEOGRAFICHE NEL DOMINIO SPAZIALE URBANO E DEL PATRIMONIO COSTRUITO

A REVIEW OF SPATIAL ONTOLOGIES FOR URBAN AND ARCHITECTURAL HERITAGE DOMAIN

Elisabetta Colucci^a *, Antonia Spanò^b

^a Dipartimento di Ingegneria dell'Ambiente, del Territorio e delle Infrastrutture (DIATI), Politecnico di Torino Corso Duca degli Abruzzi 24, 10129 Torino (Italia), elisabetta.colucci@polito.it

^b Dipartimento di Architettura e Design (DAD), Politecnico di Torino Viale Pier Andrea Mattioli 39, 10125 Torino (Italia), antonia.spano@polito.it

*corresponding author

TEMA: Artificial Intelligence in Geomatics for Cultural Heritage | **CAMPO DI RICERCA:** Ontologie

PAROLE CHIAVE: ontologie spaziali, patrimonio costruito urbano, standards, interoperabilità

KEY WORDS: spatial ontologies, urban and architectural heritage, built heritage, spatial standards, interoperability

RIASSUNTO

Il presente lavoro ha lo scopo di presentare un'articolata ricognizione della letteratura passata e attuale, inerente le ontologie informatiche, con particolare attenzione a quelle connesse al dominio spaziale urbano e del patrimonio culturale architettonico costruito. Da un punto di vista generale, il documento offre una descrizione di definizioni, classificazioni e approcci e metodi per la creazione di ontologie nel campo del Web Semantico, dalle prime nozioni emerse negli anni Novanta alla loro evoluzione negli ultimi due decenni. Nel settore del patrimonio costruito, l'interesse, il bisogno di conoscenza e l'uso di ontologie sono cresciuti per rispondere alle necessità di condivisione e scambio di dati spaziali e alla crescente adozione di infrastrutture geografiche. Queste stanno infatti riscontrando problemi di interoperabilità tecnica, geometrica e semantica nell'integrazione di database geografici e urbani multi-scala che richiedono pertanto l'adozione di standard condivisi e linguaggi comuni.

ABSTRACT

This paper aims to report a review in the past literature about computer science ontologies, with a special view on the ones connected to the spatial domain of urban data and built heritage. From a general point of view the paper offers a write up of definitions, classifications and design approaches and methods for ontologies as they have emerged since the nineties and evolved in the last two decades. In the built heritage domain, the interest, the need of knowledge and the use of ontologies have grown to face the wide exchange of digital spatial data and the extensive adoption of spatial data infrastructures, which faced problems related to the interoperability among spatial databases, and specifically the integration of geographical and urban databases involving the adoption of standards.

1. INTRODUZIONE

Il presente lavoro ha come obiettivo una revisione della letteratura su ontologie spaziali e geografiche e, nello specifico, nel dominio del patrimonio costruito urbano e architettonico.

Le innovazioni tecnologiche per la comunicazione che si sono sviluppate a partire dai primi anni 2000, hanno portato studiosi e ricercatori a soffermarsi in modo diverso sul contenuto delle informazioni, anche su quelle spaziali, che in particolar modo interessano la Geomatica.

Questo interesse è inoltre connesso alla rivoluzione inter e multidisciplinare riguardante l'utilizzo e le possibili applicazioni di queste informazioni: "per essere comprensibili e condivisibili, i modelli della conoscenza devono combinare precisione semantica con efficacia della trasparenza cognitiva, in quanto incorporano paradigmi di modellazione sempre più sofisticati ed eterogenei. La recente esplosione di interesse per le ontologie

denota fundamentalmente questa tendenza" (Guarino & Musen 2005).

Come affermato da Guarino, all'inizio degli anni 2000, l'interesse per le ontologie, al di fuori della loro accezione filosofica, rispecchia la necessità di gestire e condividere informazioni, in diversi domini e aree di applicazioni, tra diverse discipline e attraverso l'utilizzo di un linguaggio e una semantica comune. Questo interesse è particolarmente legato alla nascita e alla diffusione del Web Semantico¹, basato sull'accesso a contenuti comuni, condivisibili e interoperabili attraverso il web. Definizione da tempo condivisa di ontologia nel campo dell'informatica è la "descrizione esplicita formale di concetti in un dominio di interesse all'interno del quale vi sono classi, proprietà e relazioni" (Guarino, 2009). Un'ontologia è pertanto un oggetto informativo o un manufatto computazionale, è una "specificazione formale ed esplicita di una concettualizzazione condivisa" (Studer et al., 1998). Le ontologie sono considerate

¹ Un'estensione del World Wide Web attraverso standards. Questo mira a rendere i dati Internet "machine-readable" (XML and Semantic Web W3C Standards Timeline, 2012-02-04).

modelli concettuali in grado di formalizzare la conoscenza di un particolare dominio di studio, esse sono utili per definire concetti e relative informazioni a livello multidisciplinare.

Anche nel campo delle informazioni spaziali, i problemi di interazione tra informazioni diverse e complementari nella sfera del patrimonio urbano e costruito possono trovare supporto nella definizione di sistemi digitali grazie allo sviluppo di rappresentazioni formali e concettuali come le ontologie. Esse possono infatti consentire il controllo digitale e l'integrazione delle informazioni al fine di condividere una struttura comune tra i diversi attori coinvolti nei processi di trasformazione e sviluppo della città e del patrimonio urbano (cittadini, scienziati, ricercatori, politici, amministrazioni locali, urbanisti, architetti, ecc.). Per le ragioni sopracitate e per il fatto che le informazioni spaziali a diversa scala sono oggi ad un punto di svolta nel campo dei sistemi di gestione delle informazioni, è sempre più necessaria l'armonizzazione dei dati e delle informazioni in strutture condivise e con formati di dati comuni. A questo proposito, nello scenario del recente incontro tra l'ambiente GIS (Geographic Information Science), per la gestione delle informazioni geografiche e spaziali a scala territoriale e urbana, e i sistemi BIM (Building Information Modelling), a scala dell'edificio e molto recentemente delle infrastrutture e della città ("Urban BIM") vi sono ancora problemi di interoperabilità e numerosi ostacoli tecnici da superare per la creazione di database univoci in grado di rappresentare informazioni sia a livello urbano sia edilizio o architettonico.

Il presente studio si colloca nello scenario sopra descritto e mira a descrivere l'evoluzione delle ontologie e le loro diverse classificazioni, fino a giungere alla definizione di ontologie nel campo del web semantico e dell'intelligenza artificiale.

I modelli di classificazione delle ontologie e i relativi scopi sono stati presi in considerazione sintetizzando i metodi e gli approcci utilizzati per la loro costruzione.

Sono qui presentati, inoltre, alcuni esempi di ontologie geografiche e relative al dominio delle informazioni spaziali urbane e del patrimonio culturale e costruito.

Il lavoro di ricognizione mira a gettare le basi per future possibili integrazioni e armonizzazioni delle conoscenze e degli standards ad oggi esistenti. Chiarire le diverse connotazioni e tipologie di ontologie spaziali è infatti fondamentale per una possibile applicazione pratica in campo urbano a diverse scale di rappresentazione.

2. EVOLUZIONE DELLE DEFINIZIONI DI ONTOLOGIA: dalla filosofia all'intelligenza artificiale

2.1 "Ontologia" e "ontologia": filosofia e informatica

Il termine "ontologia", a seconda del contesto in cui viene applicato, assume molteplici accezioni. La differenza più grande è quella tra significato filosofico e contesto applicato all'informatica e all'ingegneria (Guarino, 1998).

In filosofia si parla di *Ontologie*, senza l'articolo e con la "o" maiuscola, mentre *un'ontologia*, con l'articolo e l'iniziale minuscola si riferisce all'area informatica e a quella dell'informazione e della conoscenza. Un chiarimento sulle possibili interpretazioni del termine "ontologia" è fornito in (Guarino & Giarretta, 1995) (Tabella 1).

<i>Ontologia</i>
1. Ontologia come disciplina filosofica
2. ontologia come concettualizzazione informale
3. ontologia come formalizzazione semantica

4. ontologia come specificazione di una "concettualizzazione"

5. ontologia come rappresentazione di un sistema concettuale attraverso una teoria logica:

- caratterizzato da specifiche proprietà formali
- caratterizzato da scopi specifici

6. ontologia come vocabolario

7. ontologia come metadati

Tabella 1. Possibili interpretazioni del termine "ontologia"

(Guarino & Giarretta, 1995).

È importante distinguere la prima interpretazione di *Ontologia* (1.) dalle altre (2.-7.). La principale distinzione riguarda la loro applicazione nella filosofia. Dal punto di vista etimologico, il termine composto "Ontologia" deriva dal greco antico ὄντος, òntos che significa "essere; ciò che è" e da λόγος, lògos, "discorso logico", quindi significa letteralmente "studio dell'essere". Ci si riferisce, invece, a *un'ontologia* nelle discipline della conoscenza e nell'informatica per identificare "una sorta di oggetto informativo o artefatto computazionale" (Guarino et al., 2009).

Le definizioni più citate di ontologia in informatica sono di seguito riportate:

1. "Specificazione esplicita di una concettualizzazione" (Gruber, 1993);
2. "Specificazione formale di una concettualizzazione condivisa" (Borst, 1997);
3. "Specificazione esplicita e formale di una concettualizzazione condivisa" (Studer, 1998).

Nonostante la definizione di Gruber del 1993 sia la più diffusa e citata, essa non tiene conto di una visione condivisa tra coloro che usano tale concettualizzazione o che sono i destinatari dei vantaggi della sua adozione; a questo proposito l'accezione di Studer del 1998 considera "un contesto piuttosto che una visione individuale" (Guarino et al., 2009).

Il presente lavoro si riferisce pertanto a questa definizione condivisa del termine, che unifica le precedenti. Da queste definizioni possiamo pertanto identificare un'ontologia come una specificazione di una *concettualizzazione* di un dominio. Una concettualizzazione è un modello astratto che rappresenta le entità di un dominio in termini di concetti, relazioni e altre primitive del modello (Paragrafo 4). Pertanto, un'ontologia comprende un vocabolario semantico rappresentativo di una realtà (o dominio) caratterizzato da un insieme di definizioni che specificano il significato dei termini e dei corrispondenti concetti che i termini identificano. Un vocabolario di un'ontologia, quindi, può essere definito attraverso diversi linguaggi (ad esempio con diagrammi che rappresentano classi, attributi e relazioni).

2.2 Semantic Web e Intelligenza Artificiale

Ad oggi, le ontologie sono sempre più riconosciute e utilizzate come componenti essenziali in molti campi della scienza dell'informazione (come l'informatica, l'intelligenza artificiale, il web semantico, l'ingegneria dei sistemi, l'ingegneria dei software, l'informatica biomedica e l'architettura dell'informazione) come una forma di rappresentazione della conoscenza del mondo o di parte di esso.

Negli ultimi anni si è assistito a un rapido aumento del numero di ontologie disponibili al pubblico. Queste ontologie non sono tutte di alta qualità e alcune hanno una portata molto limitata (Falquet et al., 2011); tuttavia, ciò dimostra come lo sviluppo di ontologie non sia più solo una risorsa di grandi progetti con finanziamenti significativi. Ciò è sicuramente dovuto a diversi fattori come la diffusione del web semantico, la disponibilità di metodologie per

le ontologie in ingegneria, la definizione di ontologie nella letteratura, la presenza di numerosi corsi e tutorials e molto altro. Nel campo dell'*intelligenza artificiale* (AI) le ontologie furono inizialmente impiegate come mezzo per concettualizzare alcune parti del reale. Il primo obiettivo fu quello di permettere al sistema software di "ragionare" sulle entità del mondo reale. Uno dei primi esempi è rappresentato dall'ontologia CYC (Cycorp) (Lenat, 1995), il progetto mirava a unire una base ontologica con la conoscenza di base. Tra i progetti nel campo dell'AI, il progetto CYC risultò rilevante per lo sviluppo del *web semantico* (una rete di dati collegati, chiamati *linked-data*) (Smith, 2003). Un altro esempio è l'ontologia EL, sviluppata per intelligenza artificiale e l'analisi del linguaggio naturale. Si tratta di un'ontologia che include tipi, idee, fatti ed eventi, e anche un motore di inferenza capace di deduzioni complesse in modo efficiente. Pertanto, nel campo dell'informatica, un'ontologia è una specificazione dei significati dei simboli in un sistema di informazioni e di ciò che si presume esista con una determinata semantica. Un'ontologia in AI specifica quali tipi di individui e classi verranno modellati, quali proprietà verranno utilizzate e fornisce alcuni assiomi che limitano l'uso di un determinato vocabolario. Il Web semantico consiste in un insieme di tecnologie standard per realizzare una rete di dati tramite un modello comune fornito alle macchine con il fine di descrivere e interrogare i dati con una corretta classificazione di termini e in uno specifico dominio. Tra i diversi linguaggi standard adottati nel web semantico (come XML, "Extensible Markup Language", URI, "Uniform Resource Identifier", RDF, "Resource Description Framework" e RDFs schema) è bene ricordare che l'OWL, "Web Ontology Language", è utilizzato per descrivere concetti, attributi ed esempi; possiede inoltre meccanismi incorporati in grado di assegnare proprietà di uguaglianza, transitività e cardinalità tra individui, classi e sottoclassi all'interno di domini definiti.

2.3 Cenni alle ontologie nel dominio del patrimonio costruito e urbano

L'interesse per le ontologie da utilizzare nel dominio urbano è stato inizialmente innescato da sfide tecnologiche legate all'interoperabilità delle banche dati urbane e territoriali e dalla necessità di interconnettere i diversi database. Dal momento che attualmente molti database urbani e del territorio sono stati caratterizzati dalla diffusione della tecnologia GIS per la loro gestione e per il conseguente governo dei territori di interesse (Falquet et al., 2011), un ulteriore aggiornamento di queste basi di dati che li rendesse facilmente disponibili e collegabili ad altre fonti di dati è parso impossibile senza apportare una profonda ristrutturazione del loro contenuto. Data l'entità e la complessità del processo, l'ingegneria ontologica è stata vista come un passo necessario per gestire sia la solidità concettuale che la continuità con le versioni precedenti delle basi di dati. Proprio tale integrazione tra differenti database costituisce la motivazione per sviluppare un'ontologia urbana e del territorio, questa necessità scaturisce inoltre dalla necessità di uniformare tutte le basi di dati agli standards internazionali (come ad esempio INSPIRE²). La progettazione di ontologie urbane è vista come una sfida stimolante a livello concettuale in quanto necessita un chiarimento dei mezzi di comunicazione e degli obiettivi tra i diversi attori coinvolti nelle attività di sviluppo e gestione urbana (ingegneri, urbanisti, costruttori, architetti, cittadini, ecc). Nel Paragrafo 5.2 sono presentati alcuni esempi di ontologie nel dominio del patrimonio urbano, architettonico e culturale.

² <https://inspire.ec.europa.eu/>

3. CLASSIFICAZIONI DELLE ONTOLOGIE

Le ontologie sono state utilizzate negli ultimi decenni per molteplici attività al fine di migliorare la comunicazione tra agenti (umani o software) e per riutilizzare e aggiornare modelli di dati o schemi di conoscenza. Queste attività sono caratterizzate da problemi di interoperabilità e possono essere applicate in domini diversi. Di conseguenza, con l'evolversi delle ontologie in diversi scenari sono stati proposti diversi tipi di ontologie.

Una recente revisione sulle classificazioni di ontologie esistenti è riportata in (Kokla & Guilbert, 2020). La ricerca spiega che tradizionalmente le ontologie sono classificate secondo due criteri: il loro livello di formalità e il loro livello di generalità. Secondo la prima, le ontologie sono considerate dall'informale al semi-formale e formale (Gruber, 2004). Secondo la generalità, e come descritto in seguito secondo "lo scopo", si distinguono quattro tipi di ontologie: ontologie di *primo livello* (*top-level*, *upper-level* o *foundation*), di *dominio* (*domain*), di *attività* (*task*) e di *applicazione* (*application*) (Guarino, 1998).

Le ontologie di primo livello definiscono concetti generali e fondamentali, come entità, proprietà e relazioni. Questi concetti sono indipendenti dal dominio di interesse. Le ontologie di dominio definiscono concetti relativi ad un determinato punto di vista in una certa area di interesse, mentre le ontologie "task" definiscono concetti relativi ad una specifica attività. Le ontologie geografiche possono essere considerate ontologie di dominio. Tuttavia, anche le ontologie di primo livello sono rilevanti per la conoscenza geo-spaziale in quanto definiscono nozioni generali come spazio, tempo, confini e processi. Le ontologie applicative, d'altra parte, definiscono concetti relativi a un determinato dominio e attività; queste possono specializzare entrambe le ontologie.

In Falquet et al. (2011) sono riportate ulteriori classificazioni che considerano una suddivisione basata su tipologia del linguaggio e su obiettivo e scopo. I paragrafi 3.1 e 3.2 riportano le suddette classificazioni.

3.1 Classificazioni di ontologie in base a semantica e linguaggio

Dalle meno formali alle più formali, le ontologie possono essere suddivise in quattro tipologie (Figura 1):

1. *Ontologie di informazione:*

diagrammi e schemi utilizzati dall'uomo per trasmettere progetti e idee. Queste sono sintetiche, schematiche, facilmente editabili e focalizzate su concetti esempi e relazioni.

2. *Ontologie linguistiche/terminologiche:*

glossari, dizionari, vocabolari, tassonomie, tesauri e database del lessico. Sono focalizzate su termini, relazioni e gerarchie. Il loro scopo è definire i termini più appropriati per rappresentare un concetto evitando ambiguità. Questo tipo di ontologie possono essere espresse in linguaggio RDF, "Resource Description Framework", utilizzato per descrivere i metadati web. RDF è implementato in XML ed è composto da triple (soggetto, predicato e oggetto).

3. *Ontologie software* (o ontologie basate sull'implementazione del software):

vengono utilizzate per fornire uno schema concettuale utile per la memorizzazione o la manipolazione dei dati. Nelle attività di sviluppo software vengono applicate al fine di garantire la coerenza dei dati. Un esempio di linguaggio per questo tipo di ontologie è UML (Unified Modeling Language), un linguaggio di modellazione standardizzato costituito da una serie integrata

di diagrammi, sviluppata per aiutare gli sviluppatori di sistemi e software a specificare, visualizzare, costruire e documentare gli artefatti dei sistemi software. Un esempio di ontologia software è rappresentato dallo standard IFC (Industry Foundation Classes).

4. *Ontologie formali:*

queste ontologie richiedono una semantica chiara e precise ragioni per distinguere i concetti. Nel Web semantico sono definiti con software di editor di ontologie come Protégé (un editor open source gratuito per la creazione di ontologie creato dalla Stanford University) e possono essere espresse in OWL.

\\classificazione delle ontologie

A. BASATE SUL LINGUAGGIO UTILIZZATO



Figura 1. Classificazione di ontologie in base al linguaggio utilizzato (Lassila & McGuinness, 2001).

3.2 Classificazioni di ontologie basate su obiettivo e scopo

La seconda classificazione si basa sullo scopo dell'ontologia. Come accennato precedentemente, la Figura 2 mostra come sia possibile collegare tra loro le diverse tipologie in base all'obiettivo.

1. *Ontologie locali o di applicazione:*

le ontologie locali o applicative sono specializzazioni di ontologie di dominio in cui potrebbe non esserci condivisione delle conoscenze. Questo tipo di ontologia rappresenta il modello di un dominio basato su un singolo punto di vista di un utente o di uno sviluppatore. In (Fonseca et al., 2000) le ontologie locali vengono presentate come la combinazione delle ontologie di dominio e le ontologie di attività.

2. *Ontologie di dominio e di attività:*

come sopraccitato, sono applicabili a un dominio con uno specifico punto di vista (definendo come un gruppo circoscritto di utenti concettualizza e visualizza uno specifico fenomeno). L'ontologia dell'attività (task) contiene le conoscenze per realizzare un'attività, mentre l'ontologia di dominio descrive le conoscenze in cui l'attività viene applicata.

3. *Ontologie di riferimento di base (core):*

è uno standard utilizzato da diversi gruppi di utenti; è quindi connessa a un dominio, ma integra differenti punti di vista del gruppo. È il risultato dell'integrazione tra diverse ontologie (Fonseca et al., 2002). Un esempio rappresentativo di questa ontologia è CityGML (Paragrafo 5.1), uno standard per la rappresentazione, la memorizzazione e lo scambio di modelli 3D di città. CityGML definisce le classi e le relazioni per rappresentare gli oggetti del database topografico più rilevanti rispetto alle loro proprietà geometriche, topologiche, semantiche e di aspetto. Pertanto, CityGML potrebbe essere considerato un'ontologia di riferimento di base relativa al modello di città 3D, considerando i fini della visualizzazione.

4. *Ontologie generali:*

non sono riferite a un dominio particolare, ma descrivono una vasta area di conoscenza.

5. *Ontologie di primo livello:*

sono ontologie generiche adatte a diversi domini, esse definiscono nozioni di base, oggetti, relazioni, eventi e processi.

Le ontologie di base possono essere paragonate ai modelli concettuali (Fonseca et al., 2003) per la progettazione e la strutturazione di database spaziali. Un esempio è il Geography Markup Language (GML), uno standard di codifica OpenGIS per la rappresentazione, la memorizzazione e lo scambio di informazioni geografiche.

\\classificazione delle ontologie

B. BASATE SU OBIETTIVO E SCOPO

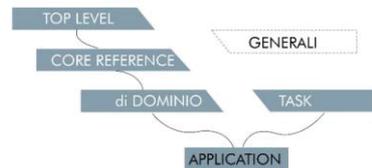


Figura 2. Classificazione delle ontologie basata sul loro scopo (Schema rielaborato da Falquet et al, 2011).

4. APPROCCI E METODI PER LA DEFINIZIONE DI ONTOLOGIE

Prima di descrivere i diversi approcci e le attività per costruire un'ontologia è importante chiarire il significato di *concettualizzazione* e *specificazione* nel quadro della creazione di ontologie.

Come schematizzato nella Figura 3, il passaggio dal mondo reale al modello esterno e infine a quello logico consiste in un processo complesso e caratterizzato da molte fasi di "modellazione dei dati" (*data modelling*). Il primo processo prevede la schematizzazione e la formalizzazione partendo dall'analisi della realtà percepita all'interno di un dominio da parte di alcuni soggetti per raggiungere un formato comprensibile dalle macchine, così da poterlo poi implementare nelle applicazioni informatiche (Laurini, Thompson, 1992).



Figura 3. Le ontologie: dal mondo reale al modello esterno (Laurini & Thompson, 1992).

Nello studio presentato in Genesereth & Nilsson (1987) la *concettualizzazione* viene descritta come "una visione astratta e semplificata del mondo che desideriamo rappresentare per qualche scopo (gli oggetti, i concetti e altre entità che si presume esistano in alcune aree di interesse e le relazioni che intrattengono tra loro). Ogni conoscenza di base o sistema basato sulla conoscenza è caratterizzato da una concettualizzazione".

Dieci anni dopo (Guarino, 1998) afferma che la nozione di Genesereth e Nilsson si riferiva a normali relazioni matematiche, bisogna invece considerare relazioni intenzionali, chiamate *relazioni concettuali*. In seguito alla definizione di queste è possibile chiarire il ruolo dell'ontologia.

In Guarino et al. (2009) vengono fornite molteplici definizioni del termine, ma la più significativa è quella in cui una

concettualizzazione è una rappresentazione matematica: una tupla (D, R) dove D è un insieme chiamato universo del discorso e R è un insieme di relazioni esistenti in D ; ogni elemento di R è una relazione che riflette un dominio nel mondo compresi gli oggetti di D (Figura 4).

È necessario inoltre definire un linguaggio (L); è importante sottolineare che un'ontologia dipende da esso, mentre una concettualizzazione è indipendente dal linguaggio. Il linguaggio è essenziale per definire e comunicare la descrizione del mondo reale. Può essere informale (linguaggio naturale, linguaggio grafico, icone ...) o formale (linguaggio logico, linguaggio matematico, linguaggio di programmazione ...). La Figura 5 mostra i diversi approcci con livelli di formalizzazione crescenti (Guarino et al., 2009).

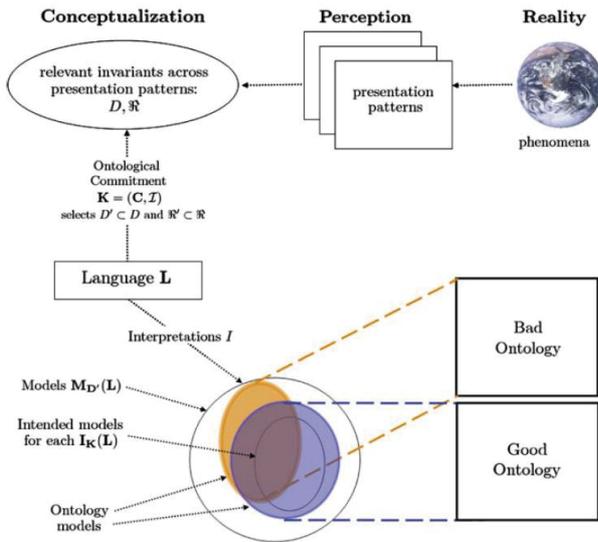


Figura 4. Le relazioni tra i fenomeni che si verificano nella realtà, la loro percezione (in tempi diversi), la loro concettualizzazione astratta, il linguaggio usato per parlare di tale concettualizzazione, i modelli previsti e un'ontologia (Guarino et al., 2009).

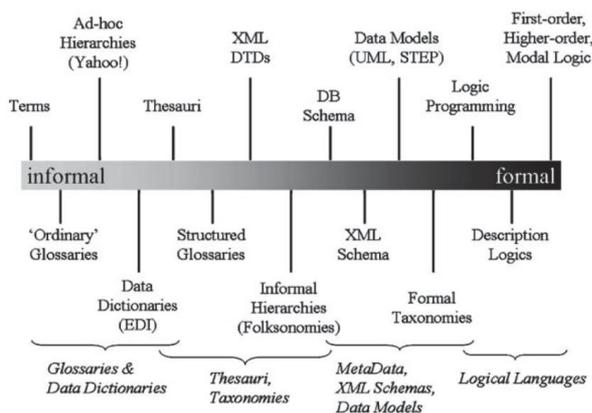


Figura 5. Approcci differenti al linguaggio. In genere, i linguaggi logici sono ammissibili per la specifica formale ed esplicita, e, quindi, per le ontologie (Guarino et al., 2009).

Negli anni sono stati proposti differenti approcci metodologici al fine di progettare un'ontologia. I più completi sono *Methontology* (Gomez-Perez et al. 2004) e *On-to-knowledge* (Sure et al. 2003). *Methontology* enfatizza il riutilizzo di domini esistenti e di ontologie di primo livello e propone di utilizzare, ai fini della formalizzazione, una serie di rappresentazioni intermedie che

possono successivamente essere automaticamente trasformate in diversi linguaggi formali. Questo metodo si basa su molteplici passaggi: specificazione del dominio di interesse, acquisizione della conoscenza, concettualizzazione (che può comprendere la creazione del glossario), formalizzazione e integrazione, implementazione in un linguaggio formale, valutazione dell'ontologia, creazione di documentazione e linee guida e manutenzione. In Figura 6 sono sintetizzati i metodi e le diverse attività da svolgere per generare un'ontologia.

La metodologia *On-to-knowledge Methodology* (OTKM) è stata sviluppata per l'introduzione di applicazioni di gestione della conoscenza nelle imprese. In particolare, il processo comprende cinque diverse fasi: lo studio di fattibilità, il kickoff, il perfezionamento, la valutazione dell'applicazione e l'evoluzione (Gomez-Perez et al., 2004).

\\metodologie e attività per costruire un'ontologia

1. specificazione
2. acquisizione conoscenza
3. concettualizzazione (modello concettuale)
4. integrazione
5. implementazione con linguaggio formale
6. validazione
7. documentazione per l'utilizzo e il riuso



3. concettualizzazione >>> approcci per la costruzione della tassonomia



Figura 6. Metodologie e attività per la creazione di un'ontologia (Gandon, 2002).

La fase di concettualizzazione è composta da diversi steps, uno dei quali è la costruzione della *tassonomia* per la quale è possibile adottare diversi approcci: top-down, bottom-up e middle-out (Gandon, 2002) (Figura 6).

L'approccio top-down parte dal concetto più generico e costruisce una struttura per specializzazione; mira a utilizzare le conoscenze esistenti (standards, ontologie, modelli concettuali, dizionari, documenti, ecc.) per definire una formalizzazione semantica di concetti specifici. Un approccio top-down viene adottato, inoltre, al fine di estendere un'ontologia integrando conoscenze già strutturate" (Kokla, et al., 2018).

L'approccio Bottom-Up, al contrario, parte dai concetti più specifici e costruisce una struttura mediante generalizzazione; l'ontologia si costruisce determinando i concetti di basso livello tassonomico e generalizzandoli. Questo approccio è incline a fornire ontologie su misura e specifiche con concetti dettagliati. Il metodo riguarda quindi l'arricchimento dell'ontologia con esempi e casi studio.

L'approccio Middle-Out, infine, identifica concetti centrali in ogni dominio identificato; i concetti chiave sono definiti e in seguito generalizzati e specializzati per popolare l'ontologia (Gandon, 2002).

Parallelamente, è bene citare la più recente e completa metodologia per la costituzione di un'ontologia nel campo del web semantico, riportata in Noy et al. (2001). Dopo aver chiarito cosa sono e perché sviluppare ontologie, il sopracitato contributo chiarisce i passaggi per il loro sviluppo. Gli autori presentano un

approccio iterativo, composto da diversi steps: determinare il dominio e lo scopo dell'ontologia (definendo le domande alla quale essa deve rispondere), considerare il riuso di ontologie esistenti, enumerare i termini e i concetti fondamentali, definire le classi e le gerarchie (secondo gli approcci sopradescritti top-down, bottom-up o misto), definire proprietà e cardinalità e, infine, validare l'ontologia con esempi.

Infine, operativamente, i metodi maggiormente usati per costruire ontologie nel dominio urbano sono l'"ontology matching", che permette di trovare corrispondenze semantiche tra le entità di ontologie diverse e l'"ontology alignment", ovvero un multi-matching che identifica corrispondenze tra più ontologie. Degli esempi applicativi sono riportati in: (Fonseca et al., 2002), (McGuinness et al. 2000) e (An et al., 2005).

5. ONTOLOGIE GEOGRAFICHE E SPAZIALI

Un'ontologia geografica definisce i concetti generali degli oggetti nello spazio e le loro relazioni in un particolare dominio. Etimologicamente, geografia significa descrizione della Terra, mentre ontologia si riferisce a ciò che esiste. Quindi, ontologia geografica sta a significare 'la descrizione delle cose esistenti sulla Terra', cioè delle sue caratteristiche geografiche. Per decenni, le ontologie sono state utilizzate nel campo delle tecnologie dell'informazione - per descrivere la conoscenza in un dominio - come reti semantiche, in particolare per l'interoperabilità di database e per la descrizione della conoscenza nell'intelligenza artificiale (Laurini et al., 2016).

In passato, le ontologie geografiche descrivevano oggetti geografici con relazioni convenzionali; ciò non era però sufficiente per descrivere lo spazio, e quindi a questo proposito è stato necessario specificare il tipo di geometria, le sue caratteristiche e il linguaggio adottato. Ad oggi, le ontologie spaziali integrano rappresentazioni dello spazio e relazioni, ovvero non rappresentano solo gli oggetti nello spazio, ma anche le relazioni topologiche che intercorrono tra questi.

In questo scenario sono state sviluppate molteplici ricerche nel campo delle ontologie geografiche e spaziali.

Nel 2002 (Fonseca et al., 2002) viene proposto l'impiego di ontologie per l'integrazione di piattaforme GIS. L'obiettivo consisteva nella definizione di un GIS in grado di integrare informazioni geografiche e semantiche. A tal fine, è stato sviluppato un modello concettuale per i dati geografici e la loro rappresentazione digitale.

Come affermato da Yeung & Hall (2007) e ampiamente discusso da Fonseca et al. (2002, 2003), il processo di progettazione e documentazione dell'ontologia è simile alla pianificazione del modello concettuale durante la creazione di database. Entrambi i processi mirano a identificare e definire caratteristiche del mondo reale e relazioni tra le entità. Tuttavia, nonostante questa similitudine, i prodotti finali non sono gli stessi. Lo scopo di uno schema concettuale è quello di descrivere la struttura del database ad un alto livello di astrazione; un'ontologia rappresenta, invece, la definizione delle relazioni semantiche e del vocabolario dei termini utilizzati per rappresentare i dati. Pertanto, non esiste necessariamente una corrispondenza diretta tra la struttura di un'ontologia e la struttura del database.

Proseguendo ancora l'analisi della letteratura su questo tema, Tomai et al. (2004; 2005a; 2005b) presentano diverse metodologie per la progettazione e l'implementazione di ontologie geografiche. La ricerca del 2004 si è focalizzata sull'analisi degli elementi delle ontologie come concetti, relazioni e assiomi al fine di sviluppare un approccio sistematico per la progettazione di ontologie geografiche. Ulteriori studi, si sono poi focalizzati sull'implementazione dell'ontologia tramite

strumenti web-based in linguaggio OWL. Infine, il lavoro del 2005 (b), esamina come la nozione di "contesto formalizzato" possa essere incorporata in un'ontologia geografica selezionando fonti di conoscenza e informazioni eterogenee.

Il lavoro di Kavouras et al. (2005) introduce una metodologia in grado di confrontare diverse categorie di ontologie geografiche. La ricerca mira a estrarre informazioni semantiche da definizioni e testi, descrivendo proprietà e relazioni semantiche per le entità o classi geografiche.

Un altro esempio di studio sulle ontologie nel campo delle informazioni geografiche è rappresentato dalla tesi di Lutz et al. (2006), una rassegna della letteratura sulle ricerche che mirano a trovare una risposta alla domanda: "come è possibile migliorare la creazione di dati geografici e servizi di informazione nelle infrastrutture di dati spaziali attraverso l'adozione di ontologie?". Ancora, lo studio di Chaves et al. (2007) propone una nuova versione del Geographic Knowledge Base (GKB): un ambiente in grado di integrare dati geografici e generare ontologie evitando la duplicazione dei dati e consentendo il riutilizzo della conoscenza.

In definitiva, in numerosi studi (Chaves et al., 2007; FT Fonseca et al., 2002; Kavouras et al., 2005; Kokla et al., 2018; Lutz et al., 2006; Tomai et al., 2004; Tomai & Spanaki, 2005) vengono presentate molteplici metodologie con l'obiettivo di arricchire e popolare ontologie geo-spaziali per estrarre informazioni semantiche.

Come è possibile comprendere dai precedenti esempi della letteratura, le ontologie geografiche sono utili e adatte in molti campi di applicazione della scienza della geo-informazione per scopi differenti: per affrontare efficacemente l'interoperabilità, per connettere basi di dati esistenti ecc.

Si può dire quindi che nello scenario della letteratura disponibile sia inclusa la possibilità di considerare le ontologie come strumento per garantire lo scambio e l'uso dei dati nel campo dell'informazione geografica (GI), in grado perciò di aiutare nei processi decisionali e di gestione delle risorse in diverse aree applicative.

Dall'evoluzione del GIS al concetto più ampio di infrastruttura dei dati territoriali (SDI), l'uso delle ontologie è stato da subito considerato di fondamentale importanza e rappresenta ad oggi una sfida per la diffusione e l'interoperabilità nell'utilizzo di piattaforme e sistemi geografici (SDIs).

Secondo il Cookbook della Global Spatial Data Infrastructure Association (Nebert, 2004), "il termine Spatial Data Infrastructure viene spesso utilizzato per indicare la raccolta di tecnologie, politiche e disposizioni istituzionali che facilitano la disponibilità e l'accesso ai dati spaziali". A questo proposito, le ontologie sono necessarie per l'integrazione e l'armonizzazione dei dati cartografici, che derivano da differenti fonti (regionali, nazionali e internazionali).

Le problematiche relative all'applicazione e all'adozione di ontologie nel campo dell'informazione geografica e alla condivisione delle informazioni spaziali mediante SDI sono legate ai numerosi formati in cui i dati sono distribuiti ed alla difficoltà di costruire un'omogenea e condivisa documentazione dei metadati. Inoltre, un altro significativo e persistente ostacolo all'interoperabilità è l'eterogeneità in termini di semantica, sintassi, topologia e geometria.

5.1 Standards per l'interoperabilità geografica e la rappresentazione dei beni culturali

Gli standard esistenti per la rappresentazione delle informazioni che concorrono alla conoscenza del patrimonio architettonico, urbano e architettonico provengono da svariati settori. Il presente paragrafo presenta gli standards più utilizzati e diffusi nell'ambito

del patrimonio edilizio e i principali standards adottati oggi nel campo dell'informazione geografica.

L'ontologia di base (core) per la rappresentazione del patrimonio culturale è il modello concettuale di riferimento CIDOC (CIDOC-CRM, Conceptual Reference Model), sviluppato dal Comitato Internazionale per la DOCUMENTAZIONE (CIDOC) (Doerr et al., 2002, 2007).

Durante gli anni sono state sviluppate alcune estensioni al fine di migliorare l'applicazione del CIDOC-CRM a diverse tipologie di patrimonio costruito e culturale, un esempio è il "Monument Damage Information System" (MONDIS) (Blaško et al., 2012; Cacciotti et al., 2013).

Oltre a questo, vi sono altri esempi di estensioni dell'ontologia "core" per beni culturali e sono disponibili differenti applicazioni e attività.

Il CRMgeo (Hiebel et al., 2015) integra le proprietà spazio-temporali delle entità, e a tale scopo, integra il CIDOC CRM con lo standard OGC GeoSPARQL. Quest'ultimo mira a rappresentare e interrogare i dati geospaziali nel web semantico, definendo con la sua struttura le classi geografiche *Spatial Object, Feature and Geometry*.

Il CRMba (Ronzino et al., 2016) propone di modellare la complessità e delinea i metadati per la documentazione del patrimonio costruito nella prospettiva della buildings archeology; il CRMarcheo (Doerr et al., 2020) supporta il processo di scavo archeologico e tutti i processi e le attività ad esso correlate e, infine, il CRMsci (Doerr, 2018) è un'ontologia formale destinata ad essere utilizzata come schema globale per l'integrazione di dati e metadati per aspetti inerenti l'osservazione scientifica.

Nel quadro della classificazione del patrimonio, è bene considerare inoltre anche vocabolari e classificazioni; il Getty Institute Vocabularies propone una gerarchia di termini, con annesse definizioni, relativi al dominio del patrimonio culturale, con l'obiettivo di classificare opere d'arte e d'architettura, nomi di artisti o architetti e categorie geografiche. Uno di questi vocabolari è il Thesaurus di Arte e Architettura (AAT).

Per quanto concerne lo scenario italiano, possiamo invece considerare alcuni standard e ontologie sviluppati dal "Ministero per i Beni, le Attività Culturali e il Turismo" (MIBACT). L'Istituto Centrale Italiano per il Catalogo e la Documentazione (ICCD - MIBACT) ha sviluppato una classificazione finalizzata alla documentazione e alla conservazione dei beni culturali; questa è implementata nella piattaforma SIGECweb e sono in corso alcune ricerche per uniformare la struttura all'ontologia CIDOC-CRM. Per garantire l'uso pratico di questa classificazione e la sua interoperabilità, sono state progettate alcune ontologie specifiche. Il primo esempio è ArCo³, il Knowledge Graph per il patrimonio culturale italiano, composto da 7 vocabolari. In secondo luogo, l'ontologia Cultural-ON⁴ mira a modellare i dati che caratterizzano i luoghi culturali.

Nel campo dell'informazione geografica, come noto, esistono numerosi standards riconosciuti e adottati a livello internazionale, dei quali si fornisce un breve richiamo nel seguito.

CityGML, pubblicato dall'Open Geospatial Consortium (OGC), è probabilmente il modello di dati standard più riconosciuto a livello internazionale per rappresentare informazioni 3D multi-scala per edifici e città. Sono stati inoltre sviluppate alcune estensioni per differenti domini denominate ADE (Application Domain Extensions) (Biljecki et al., 2018).

³ <http://wit.istc.cnr.it/arco>

INSPIRE (Interoperable Spatial Data Infrastructure in Europe) è invece la direttiva europea (pubblicata nel 2007 e adottata obbligatoriamente in tutti i paesi europei nel 2020) nata con il fine di creare un'infrastruttura di dati territoriali interoperabile in Europa. Lo standard mira alla rappresentazione di dati transfrontalieri omogenei a supporto di analisi territoriali. Nella parte del modello di dati è inclusa l'entità "edificio" (building), e alcune proprietà sono compatibili con il modello CityGML.

Per quanto riguarda l'architettura, l'ingegneria e l'edilizia, nonché la gestione di risorse e impianti, l'Industry Foundation Classes (IFC) è lo standard progettato per il Building Information Model (BIM). Lo standard IFC, certificato dal building SMART, offre una visione basata su oggetti del modello, comprendente le geometrie 3D e 2D insieme alle relazioni intercorrenti tra gli oggetti. Esso mira alla rappresentazione di informazioni relative alla progettazione di nuovi edifici, utilizzando la modellazione parametrica. Recentemente è stata proposta la serializzazione nel formato ifcOWL che è compatibile con gli standard RDF, RDFS e OWL2 del Semantic Web.

Ad oggi, l'integrazione tra tecnologie semantiche basate sul web (approcci ontologici) e l'industria AEC (Architecture, Engineering and Construction) sembra essere sempre più perseguita (Pauwels et al. 2017).

Non mancano approcci operativi che sviluppano metodologie e algoritmi per automatizzare la conversione da Building Information Models nei modelli di edifici CityGML, catturando sia l'informazione geometrica che quella semantica disponibile nei modelli BIM originari (Stouffes et al. 2018); tali procedure devono affrontare la complessità di strategie di modellazione completamente differenti, come rappresenta efficacemente la figura 7.

Partendo dal presupposto che le tecniche di rilevamento 3D basate sulla realtà abbiano contribuito alla definizione della geometria degli oggetti che confluiscono nei modelli HBIM e che le ontologie siano utili per la gestione della conoscenza, Yang et al (2019) propongono un uso combinato di queste tecnologie. In particolare, il modello HBIM viene convertito in un modello ontologico per arricchire il contenuto semantico. Ovviamente, questi esperimenti sono interessanti per orientare sempre più la raccolta e l'analisi della CH verso una conoscenza standardizzata e quindi facilmente condivisa dai molti stakeholder che contribuiscono alla conservazione del patrimonio.

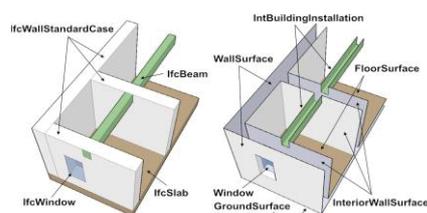


Figura 7. Differenze nel modello di un edificio tra IFC (a sinistra) e CityGML (a destra) (Nagel et al., 2009).

5.2 Esempi di ontologie nel dominio del patrimonio culturale costruito e urbano

Questa sezione riporta alcuni lavori svolti per la creazione di ontologie nel dominio urbano, presentando ricerche applicate al patrimonio architettonico e storico.

Nel dominio del patrimonio urbano e costruito sono molte le possibili applicazioni e le comunità coinvolte nei processi e nelle attività per la gestione, lo sviluppo e la valorizzazione di beni

⁴ http://dati.beniculturali.it/cultural_on/

culturali, edifici e città. In questo scenario possiamo citare ad esempio il restauro, la conservazione, la documentazione, le attività di gestione del rischio, le discipline estimative ed energetiche e molte altre.

A questo proposito, alcuni gruppi di lavoro hanno sviluppato alcuni studi che propongono ontologie (spaziali e informatiche) per la rappresentazione del costruito a diverse scale e con diversi livelli di granularità (Colucci et al., 2020).

Per quanto riguarda ad esempio l'area della *conservazione*, che consiste in uno degli ambiti più studiati, diverse ricerche hanno proposto soluzioni per l'implementazione di ontologie esistenti come il CIDOC-CRM (Tait, While, 2009; Acierno et al., 2017; Acierno et al., 2019; MONDIS, descritto in Blaško et al., 2012 e Cacciotti et al., 2013). In particolare, come descritto in seguito, Acierno et al. (2019) propongono un'ontologia per la *conservazione e la gestione* dei centri storici. Altri studi riportano invece possibili ontologie applicative per la *documentazione* del patrimonio architettonico (Hois et al., 2009). Nel campo della *valutazione energetica* di nuclei abitati ed edifici storici, è stata proposta invece un'estensione di CityGML (Egusquiza et al., 2018).

Focalizzando l'attenzione nel dominio urbano e patrimonio costruito è possibile individuare diversi casi sviluppati per la formalizzazione semantica di città, beni culturali, centri e edifici storici.

Il libro di Teller, et al. (2007), *Ontologies for urban development* e lo studio di Berdier & Roussey (2007) *Urban ontologies: The townology prototype towards case studies* descrivono due esempi concreti di ontologie nel campo di dominio urbano. Entrambi sono derivati dallo sviluppo del progetto Townology (COST action C21: Townology - Urban Ontologies for a Better Communication in Urban Civil Engineering Projects - UCE), avviato nel 2008 e terminato nel 2012. L'obiettivo principale era quello di aumentare la conoscenza nel campo dell'ingegneria civile urbana mediante ontologie per facilitare le comunicazioni tra sistemi di informazione, parti interessate e specialisti a livello europeo. A questo proposito lo studio di Teller, et al. (2007) riassume il programma dell'azione C21 dopo il seminario di novembre 2006 organizzato per affrontare le questioni emergenti nel settore. La ricerca di Berdier & Roussey (2007) descrive invece il lavoro svolto nell'ambito del progetto per studiare significati e classificazioni di ontologie per studi urbani e relative applicazioni.

Un altro esempio di ontologie nel dominio urbano è riportato nel libro di Falquet et al. (2011).

Nel quadro del patrimonio architettonico e costruito sono stati condotti alcuni tentativi per la progettazione di ontologie. Alcuni esempi sono: il lavoro di Messaoudi et al. (2015) mirato a rappresentare fenomeni di degrado nell'area archeologica; lo studio di Kokla et al. (2019) verso la costruzione di una formalizzazione semantica per (piccoli) centri storici considerando le ontologie e i modelli di dati esistenti per possibili estensioni, integrazioni e armonizzazioni.

Tra le opere recenti e significative che studiano le ontologie nel campo dell'architettura e dei beni culturali è possibile rilevare: *An ontology-based framework for conservation process* (Acierno, et al., 2017) e *Ontologie per i Centri Storici* (Acierno, M. 2019) in "Il futuro dei centri storici. Digitalizzazione e strategia conservativa" (Fiorani, D., 2019). La prima ricerca studia il patrimonio architettonico in un quadro ontologico per attività conservative considerando la possibilità di modellare l'edificio in ambiente BIM. In Acierno (2019) è stata invece proposta una metodologia per strutturare un'ontologia per valutare la

vulnerabilità dei centri storici (considerando e implementando la Carta del Rischio italiana e l'ontologia geografica GeoSPARQL – OGC).

Inoltre, è importante rilevare che una componente rilevante del progetto INCEPTION⁵ ha avuto per obiettivo lo sviluppo di una ontologia HBIM destinata alla conoscenza semantica del patrimonio costruito e ambienti connessi, utilizzando la tecnologia del *Semantic Web*. (Bonsma P. et al. 2016). La piattaforma Inception basata sull'open standard semantic web è dotata di un motore di ricerca semantico ed è studiata per costituire un archivio di modelli 3D con informazioni semantiche associate

6. CONCLUSIONI E PROSPETTIVE FUTURE

Il lavoro riporta un'indagine sullo stato dell'arte nel campo delle ontologie spaziali per possibili applicazioni nel dominio urbano, con lo scopo di analizzare la letteratura passata ed esistente al fine di pianificare gli studi futuri.

L'obiettivo è stato quello di fornire una panoramica delle classificazioni, metodologie e possibili campi di applicazione in cui le ontologie potrebbero essere adottate per risolvere problemi di interoperabilità semantica e per lo scambio di informazioni. I problemi di interazione tra informazioni diverse e complementari nel settore del patrimonio urbano, costruito e culturale possono quindi trovare supporto nella definizione di sistemi digitali grazie allo sviluppo di rappresentazioni formali e concettuali quali le ontologie. Esse possono consentire il controllo digitale delle informazioni e sono utili per condividere informazioni tra cittadini, scienziati, ricercatori, responsabili politici e altre parti interessate. Nonostante vi siano numerosi esempi di ontologie geografiche e spaziali applicate al territorio e al patrimonio costruito, la ricognizione qui riportata evidenzia la mancanza di una struttura comune che incorpori le diverse definizioni e rappresentazioni del patrimonio costruito a diverse scale e comprendente informazioni proprie di diverse comunità e campi applicativi.

Questo studio quindi, evidenziando alcuni punti critici o evidenti gap da colmare, getta le basi per un possibile sviluppo di ontologie di dominio o applicative per la valorizzazione, gestione, documentazione del patrimonio architettonico e urbano. L'armonizzazione e l'integrazione degli esempi e delle conoscenze esistenti, sopra presentate, possono infatti essere utilizzate come punto di partenza per una nuova concettualizzazione e formalizzazione semantica.

Inoltre, nell'attuale scenario caratterizzato da un'ampia disponibilità di dati spaziali 2D e 3D derivati e interrelati all'innovazione tecnologica, le ontologie di dominio potrebbero supportare diversi processi, come l'estrazione automatica di informazioni utili, il riconoscimento di edifici storici all'interno di nuclei urbani, la segmentazione automatica di elementi significativi, l'identificazione di parti di edifici e città e così via. Ancora, nel campo dell'intelligenza artificiale l'utilizzo di ontologie potrebbe supportare metodologie *machine learning* o *deep learning* aiutando a definire e identificare classi di oggetti tramite l'uso di regole semantiche predefinite

Concludendo, questo contributo rappresenta una fase preliminare di una ricerca che mira a comprendere quanto già esiste nel campo delle ontologie geo-spaziali per il patrimonio edilizio e urbano, con il fine di valutare il riutilizzo delle conoscenze precedenti per ovviamente progettare e strutturare nuove formalizzazioni. Grazie alla ricognizione delle metodologie applicabili nell'area del web semantico sarà possibile scegliere

⁵ <https://www.inception-project.eu/en>

l'approccio più adatto per l'adattamento o lo sviluppo di una struttura ontologica destinata alla definizione del dominio degli edifici e del loro contesto per lo sviluppo di diverse attività di interesse culturale e urbano.

Riferimenti

- Acierno, M., Cursi, S., Simeone, D., & Fiorani, D. (2017). Architectural heritage knowledge modelling: An ontology-based framework for conservation process. *Journal of Cultural Heritage*, 24, 124–133. <https://doi.org/10.1016/j.culher.2016.09.010>
- Acierno, M. (2019). Ontologie per i Centri Storici. In Fiorani, D. (2019). *Il futuro dei centri storici. Digitalizzazione e strategia conservativa*.
- An, Y., Borgida, A., & Mylopoulos, J. (2005). Inferring complex semantic mappings between relational tables and ontologies from simple correspondences. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 3761 LNCS, 1152–1169. https://doi.org/10.1007/11575801_15
- Berdier, C., & Roussey, C. (2007). Urban ontologies: The townology prototype towards case studies. *Studies in Computational Intelligence*, 61, 143–155. https://doi.org/10.1007/978-3-540-71976-2_13
- Biljecki, F., Kumar, K., & Nagel, C. (2018). CityGML application domain extension (ADE): overview of developments. *Open Geospatial Data, Software and Standards*, 3(1), 13.
- Blaško, M., Cacciotti, R., Křemen, P., Kouba, Z. (2012). Monument damage ontology. *Progress in Cultural Heritage Preservation*. Springer Berlin Heidelberg. pp. 221-230.
- Borst, W. (1997). Construction of Engineering Ontologies. In Centre of Telematica and Information Technology, Retrieved from <http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:C ONSTRUCTION+OF+ENGINEERING+ONTOLOGIES#1>
- Bonsma P. et al. (2016) INCEPTION Standard for Heritage BIM Models. In: Ioannides M. et al. (eds) *Digital Heritage. Progress in Cultural Heritage: Documentation, Preservation, and Protection*. EuroMed 2016. *Lecture Notes in Computer Science*, vol 10058. Springer, Cham
- Cacciotti, R., Valach, J., Kuneš, P., Čerňanský, M., Blaško, M., řemen, P. (2013). Monument damage Information System (MONDIS), An Ontological Approach to Cultural Heritage Documentation. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2(5), pp.55-60.
- Chaves, M. S., Rodrigues, C., & Silva, M. (2007). Data model for geographic ontologies generation. *XML: Aplicações e Tecnologias Associadas (XATA 2007)*, 47–58.
- Colucci, E, Kokla, M., Mostafavi, M.A., Noardo, F., Spano, A., 2020. Semantically Describing Urban Historical Buildings Across Different Levels Of Granularity. IN PRESS. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*.
- Doerr M. (2002). The CIDOC CRM: an ontological approach to semantic interoperability of metadata. *AI Magazine, Special Issue*, 24(3), pp. 75-92.
- Doerr, M., Ore, C.-E., & Stead, S. (2007). The CIDOC Conceptual Reference Model - A New Standard for Knowledge Sharing ER2007 Tutorial. Tutorials, Posters, Panels and Industrial Contributions at the 26th International Conference on Conceptual Modeling - Volume 83, 83(Er), 51–56. Retrieved from <http://dl.acm.org.acm.han.bg.pg.edu.pl/citation.cfm?id=1386963&CFID=742484481&CFTOKEN=46790097>
- Doerr M., G. Hiebel, O. Eide, (2013). CRMgeo: Linking the CIDOC CRM to geoSPARQL through a Spatiotemporal Refinement. *Tech. Rep. GR70013*, Institute of Computer Science, 2013.
- Doerr, M., Felicetti, A., Hermon, S., Hiebel, G., Kritsotaki, A., Masur, A., May, K., Ronzino, P., Schmidle, W., Theodoridou, M., Tsiafaki, D., Christaki, E., (2020). Definition of the CRMarchaeo. An Extension of CIDOC CRM to support the archaeological excavation process, version 1.5.0.
- Doerr, M., Kritsotaki, A., Rousakis, Y., Hiebel, G., Theodoridou, M., (2018). Definition of the CRMsci. An Extension of CIDOC CRM to support scientific observation, version 1.2.5.
- Egusquiza, A., Prieto, I., Izgara, J. L., Béjar, R. (2018). Multiscale urban data models for early-stage suitability assessment of energy conservation measures in historic urban areas. 87-98.
- Falquet, G., Métal, C., Teller, J., & Tweed, C. (2011). Ontologies in Urban Devopment Projects. In Springer. <https://doi.org/10.1017/CBO9781107415324.004>
- Fiorani, D. (2019). *Il futuro dei centri storici. Digitalizzazione e strategia conservativa, Percorsi. Città e architettura nel tempo*, Qasar, 9788871409252.
- Fonseca, F., Egenhofer, M., Davis, C., Borges, K., (2000). Ontologies and knowledge sharing in Urban GIS. *Comput. Environ. Urban. Syst.* 24(3), 232–251
- Fonseca, F., Egenhofer, M., Davis, C., & Câmara, G. (2002). Semantic granularity in ontology-driven geographic information systems. *Annals of Mathematics and Artificial Intelligence*, 36(1–2), 121–151. <https://doi.org/10.1023/A:1015808104769>
- Fonseca, F., Davis, C., & Camara, G. (2003). Bridging Ontologies and Conceptual Schemas in Geographic Applications Development. *Geoinformatica*, 7(4), 355–378.
- Fonseca, F., Câmara, G., Monteiro, A.M., (2006). A framework for measuring the interoperability of geo-ontologies. *Spat. Cogn. Comput* 6(4), 307–329
- Fonseca, F., (2007). The double role of ontologies in information science research. *J. Am. Soc. Inf. Sci. Technol.* 58(6), 786–793
- Gandon, F., (2002). Distributed artificial intelligence and knowledge management: Ontologies and multiagent systems for a corporate semantic web. Scientific Philosopher Doctorate Thesis in Informatics, Defended Thursday the 7th of November 2002, INRIA and University of Nice - Sophia Antipolis, Doctoral School of Sciences and Technologies of Information and Communication
- Geneşereth, M. R. & Nilsson, N. J. (1987). *Logical Foundations of Artificial Intelligence*. San Mateo, CA: Morgan Kaufmann.
- Gómez-Pérez, A. (2004). Ontology evaluation. In *Handbook on ontologies* (pp. 251-273). Springer, Berlin, Heidelberg.
- Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition*, pp. 199–220. <https://doi.org/https://doi.org/10.1006/knac.1993.1008>
- Gruber, T. 2004. Every Ontology Is a Treaty. Interview at <http://www.sigsemis.org> from Miltiadis Lytras on 2004-11-22. Available at: <http://lists.w3.org/Archives/Public/www-annotation/2004JulDec/0032.html>
- Gruber, T. (2004). Every Ontology Is a Treaty. Interview at <Http://Www.Sigsemis.Org> from Miltiadis Lytras on 2004-11-22., 1(3), 1–5. Retrieved from <ttp://lists.w3.org/Archives/Public/www-annotation/2004JulDec/0032.html>
- Guarino, N., & Giaretta, P. (1995). Ontologies and Knowledge Bases: Towards a Terminological Clarification. *Towards Very Large Knowledge Bases. Knowledge Building and Knowledge Sharing*, 1(9), 25–32. <https://doi.org/10.1006/ijhc.1995.1066>

- Guarino, N., (1998). Formal ontology in information systems: Proceedings of the first international conference (FOIS'98), June 6-8, Trento, Italy. In *Frontiers in Artificial Intelligence and Applications*; Guarino, N., Ed.; IOS Press: Amsterdam, The Netherlands, 1998; pp. 3–15, ISBN 978-90-5199-399-8.
- Guarino, N., & Musen M.A. (2005). *Applied Ontology: Focusing on Content*, *Applied ontology* 1(1):1-5.
- Guarino, N., Oberle, D., & Staab, S. (2009). What Is Ontology? *Handbook on Ontologies, International Handbooks on Information Systems*, 1–17. <https://doi.org/10.1007/978-3-540-92673-3>
- C.H. Hwang and L.K. Schubert, EL: a representation that lets you say it all. In: *Proceedings of the International Workshop on Formal Ontology: Conceptual Analysis and Knowledge Representation*, Padova, Italy, 1993, pp. 277–290.
- Hiebel, G., Doerr, M., Eide, O., & Theodoridou, M. (2015). CRMgeo: a Spatiotemporal Model. An Extension of CIDOC CRM to link the CIDOC CRM to GeoSPARQL through a Spatiotemporal Refinement. Proposal for approval by CIDOC CRM-SIG, version 1.2
- Hois, J., Bhatt, M., & Kutz, O. (2009). Modular ontologies for architectural design. *Frontiers in Artificial Intelligence and Applications*, 198(1), 66–77. <https://doi.org/10.3233/978-1-60750-047-6-66>
- Kokla, M., Kavouras, M., (2002). Extracting latent semantic relations from definitions to disambiguate geographic ontologies. In: *GIScience 2002 Abstracts, Second International Conference on Geographic Information Science*. Boulder, CO, pp. 87–90.
- Kavouras, M., Kokla, M., & Tomai, E. (2005). Comparing categories among geographic ontologies. *Computers and Geosciences*, 31(2), 145–154. <https://doi.org/10.1016/j.cageo.2004.07.010>
- Kokla, M., Mostafavi, M. A., Noardo, F., & Spanò, A. (2019). Towards building a semantic formalization of (small) historical centres. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2/W11(2/W11), 675–683. <https://doi.org/10.5194/isprs-archives-XLII-2-W11-675-2019>
- Kokla, Margarita, & Guilbert, E. (2020). A review of geospatial semantic information modeling and elicitation approaches. *ISPRS International Journal of Geo-Information*, 9(3). <https://doi.org/10.3390/ijgi9030146>
- Kokla, M., Papadias, V., & Tomai, E. (2018). Enrichment and population of a geospatial ontology for semantic information extraction. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, 42(4), 379–382. <https://doi.org/10.5194/isprs-archives-XLII-4-309-2018>
- Lassila, O., & McGuinness, D. (2001). The role of frame-based representation on the semantic web. *Linköping Electronic Articles in Computer and Information Science*, 6(5), 2001.
- Laurini, R., & Thompson, D. (1992). *Fundamentals of spatial information systems* (Vol. 37). Academic press.
- Laurini, R., Servigne, S. (2011). Potentialité du géoweb. *L'Espace géographique*, (2), pp. 109-116.
- Lutz, M., Klien, E., (2006). Ontology-based retrieval of geographic information. *IJGIS* 20(3), 233–260
- Lenat, D. B. (1995). CYC: A large-scale investment in knowledge infrastructure. *Communications of the ACM*, 38(11), 33-38.
- McGuinness, D. L., Fikes, R., Rice, J., & Wilder, S. (2000). The chimaera ontology environment. *AAAI/IAAI*, 2000, 1123-1124.
- Messaoudi, T., De Luca, L., & Véron, P. (2015). Towards an ontology for annotating degradation phenomena. 2015 Digital Heritage International Congress, Digital Heritage 2015, 379–382. <https://doi.org/10.1109/DigitalHeritage.2015.7419528>
- Nagel, C., Stadler, A. and Kolbe, T.H. (2009), Conceptual Requirements for the Automatic Reconstruction of Building Information Models from Uninterpreted 3d Models, in T.H. Kolbe, H. Zhang and S. Zlatanova (eds.), *Academic Track of Geoweb 2009: Cityscapes*, ISPRS, 46-53.
- Nebert, D. D. (2004). *The SDI Cookbook—Developing Spatial Data Infrastructures*. Global Spatial Data Infrastructure Association, 2.
- Noy, N. F., & McGuinness, D. L. (2001). *Ontology development 101: A guide to creating your first ontology*.
- Pauwels, P.; Zhang, S.; Lee, Y.-C. (2017), Semantic web technologies in AEC industry: A literature overview, *Automation in Construction*, 73, 145-165, doi: 10.1016/j.autcon.2016.10.003.
- Ronzino, P., Niccolucci, F., Felicetti, A., Doerr, M., (2016). Definition of the CRMba. An extension of CIDOC CRM to support buildings archaeology documentation, version 1.4, December 2016 Sharing, IJCAI-95, Montreal.
- Smith, B. (2003). *Ontology*. 166, 155–166.
- Studer, R., Benjamins, V. R., & Fensel, D. (1998). *Knowledge Engineering: Principles and methods. Data and Knowledge Engineering*, 25(1–2), 161–197. [https://doi.org/10.1016/S0169-023X\(97\)00056-6](https://doi.org/10.1016/S0169-023X(97)00056-6)
- Stouffs R., Tauscher H., Biljecki F., (2018) Achieving Complete and Near-Lossless Conversion from IFC to CityGML, *ISPRS Int. J. Geo-Inf.* 2018, 7, 355; doi:10.3390/ijgi7090355.
- Sure, Y., Akkermans, H., Broekstra, J., Davies, J., Ding, Y., Duke, A., ... & Kampman, A. (2003). *On-To-Knowledge: Semantic Web-Enabled Knowledge Management*. In *Web intelligence* (pp. 277–300). Springer, Berlin, Heidelberg.
- Yeung, A. K., & Hall, G. B. (2007). *Spatial database systems: Design, implementation and project management* (Vol. 87). Springer Science & Business Media.
- Yang X, Lu Y C, Murtiyoso A, Koehl M, Grussenmeyer P, (2019) HBIM Modeling from the Surface Mesh and Its Extended Capability of Knowledge Representation, *ISPRS Int. J. Geo-Inf.*, 8, 301; doi:10.3390/ijgi8070301
- Teller, J., Lee, J. R., & Roussey, C. (Eds.). (2007). *Ontologies for urban development* (Vol. 61). Springer.
- Tait, M., & While, A. (2009). Ontology and the conservation of built heritage. *Environment and Planning D: Society and Space*, 27(4), 721-737.
- Tomai, E., & Kavouras, M. (2004). From “Onto-GeoNoesis” to “Onto-Genesis”: The design of geographic ontologies. *GeoInformatica*, 8(3), 285–302. <https://doi.org/10.1023/B:GEIN.0000034822.47211.4a>
- Tomai, E., & Kavouras, M. (2005). Context in Geographic Knowledge Representation. How the notion of formalized context can be incorporated into a geographic ontology. *Proceedings of the II International Conference & Exhibition on Geographic*
- Tomai, E., & Spanaki, M. (2005). From ontology design to ontology implementation: A web tool for building geographic ontologies. *Proceeding of AGILE 2005*. Retrieved from http://www.agile-online.org/Conference_Paper/CDs/agile_2005/papers/72_EleniTomai.pdf