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 publicy 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, oggigiorno, 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.

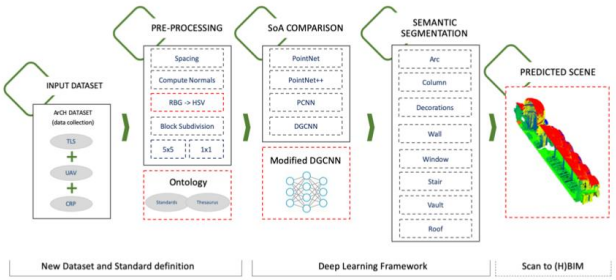

Figura 1. *Workflow* per la segmentazione semantica delle nuvole di punti.

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 volte 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.

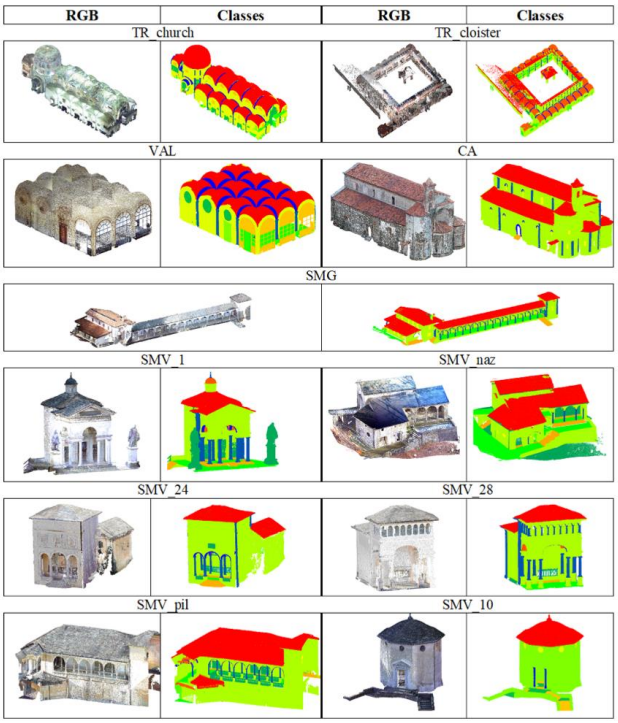

Figura 2. Scene dell'ArCH dataset utilizzate per i test. Nella colonna di sinistra le nuvole di punti RGB e in quella a destra le scene annotate. Sono state identificate 10 classi: Arco, Colonna, Porta, Pavimento, Tetto, Scale, Volta, Muro, Finestra e Decorazione. La classe Decorazione comprende tutti 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 l'*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.

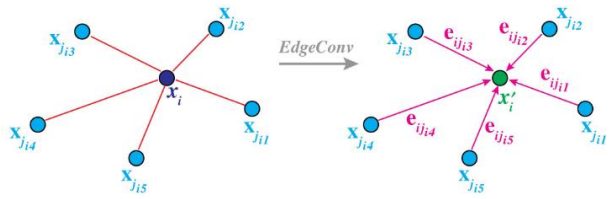

Figura 3. Modulo *EdgeConv*. L'output di EdgeConv viene calcolato aggregando le *features* del bordo associate a tutti i bordi provenienti da ciascun vertice connesso.

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.

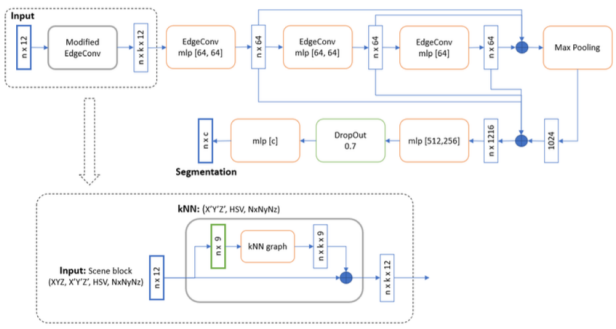

Figura 4. Illustrazione dell'architettura della DGCNN modificata.

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.

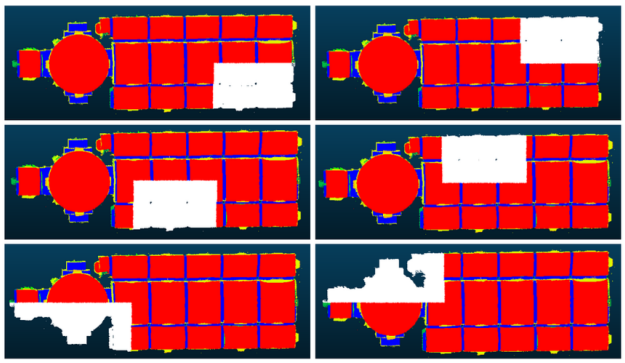

Figura 5. *6-fold Cross Validation* sulla scena TR_church. La *fold* bianca in ogni esperimento è la parte utilizzata per il test

Rete	*Feautures*	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_1score
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, F_1-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.

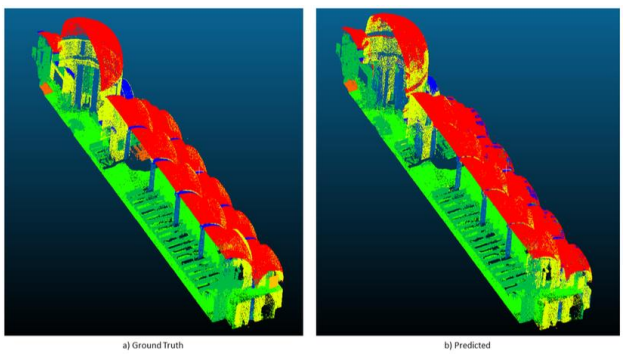

Figura 6. *Ground Truth* e nuvola di punti predetta, utilizzando la DGCNN modificata, sul lato di test di Trompone.

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 F_1score. 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 F_1score.

Rete	Test	Prec.	Rec.	F_1score
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.

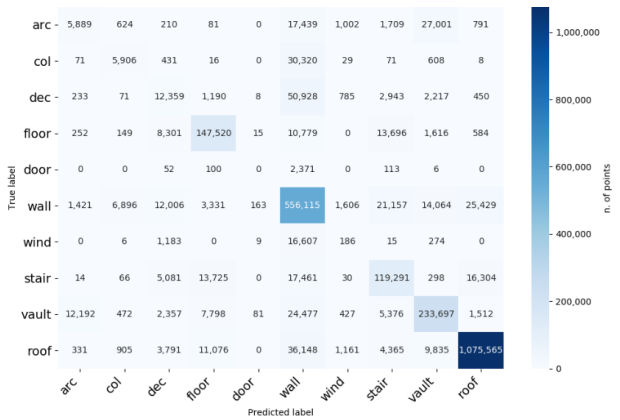

Figura 7. Matrice di confusione per l'ultimo esperimento: 9 scene per l'addestramento, 1 scena per la validazione e 1 scena per il test.

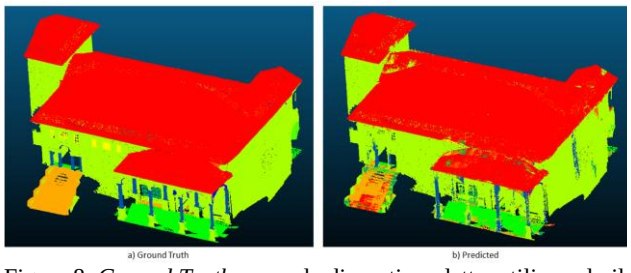

Figura 8. *Ground Truth* e nuvola di punti predetta, utilizzando il la DGCNN modificata.

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 osseva 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.

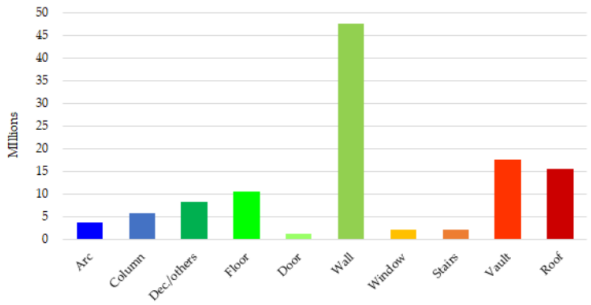

Figura 9. Numero di punti per classe

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 modellng: 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., Dininno, 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. *ISPAr*, *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. Spatial 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*.