POLITECNICO DI TORINO
Repository ISTITUZIONALE

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

Original
Un benchmark per la segmentazione semantica di nuvole di punti di beni culturali / Matrone, F., Lingua, A.M., Pierdicca, R., Malinverni, E.S., Paolanti, M., Grilli, E., Remondino, F., Muriyoso, A., Landes, T.. - In: BOLLETTINO DELLA SOCIETÀ ITALIANA DI FOTOGRAMMETRIA E TOPOGRAFIA. - ISSN 1721-971X. - ELETTRONICO. - 1(2020), pp. 10-18.

Availability:
This version is available at: 11583/2909514 since: 2021-06-27T01:16:41Z

Publisher:
SIFET

Published
DOI:

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

10 September 2026

© Autori 2021 - Articolo sottoposto a revisione

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

A BENCHMARK FOR LARGE-SCALE HERITAGE POINT CLOUD SEMANTIC SEGMENTATION

F. Matrone [1], A. Lingua [1,2], R. Pierdicca [3], E.S. Malinverni [3], M. Paolanti [4], E. Grilli [5], F. Remondino [5], A. Murtiyoso [6], T. Landes [6]

[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 *featuers* "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, *Nx, Ny, Nz.*

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 Commitee 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 BBCC, 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).

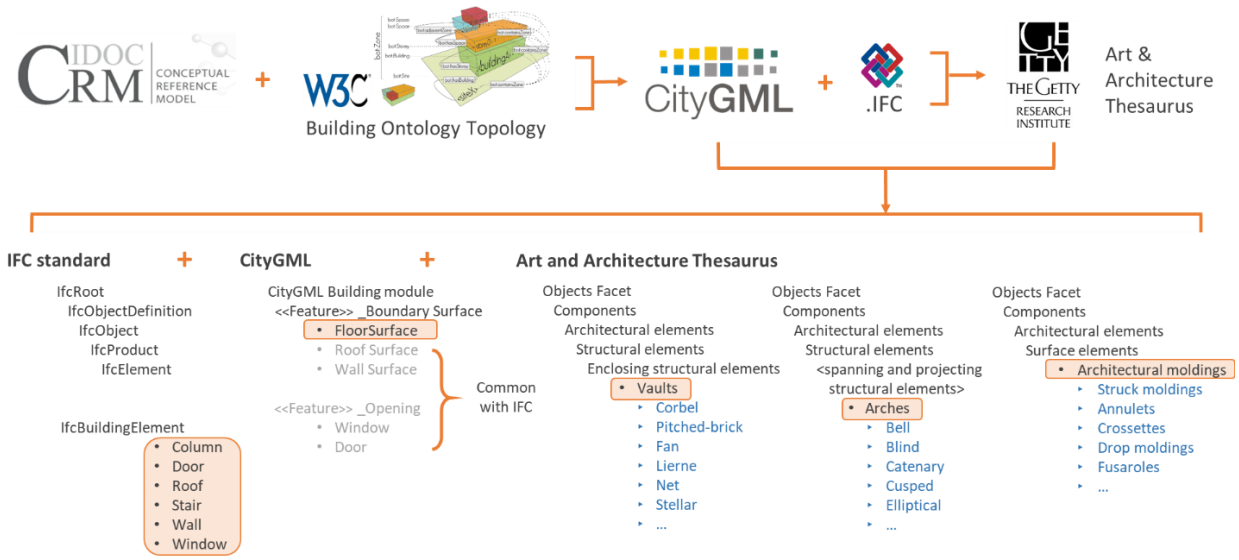

Figura 3. Classi selezionate per l'annotazione delle nuvole di punti.

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).

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

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

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

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

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

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.	F_1Score
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.

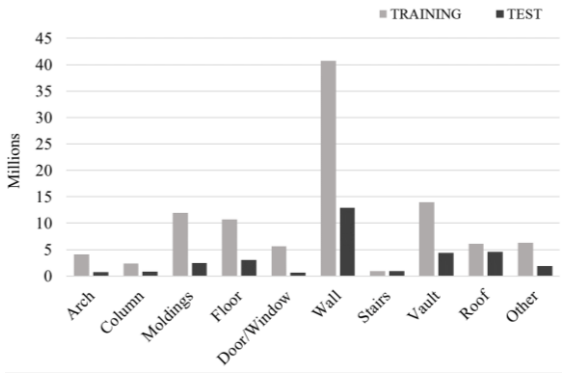

Figura 4. Numero di punti per classe.

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 nuvole di 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.

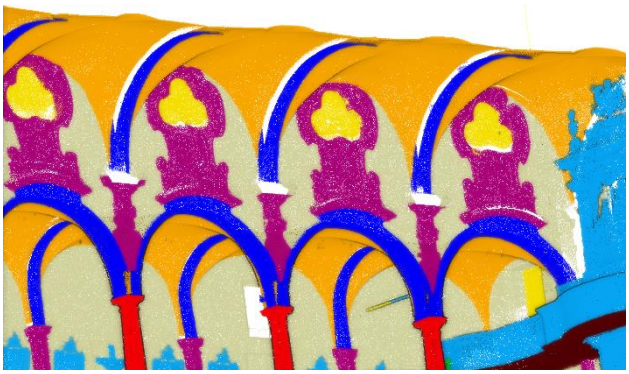

Figura 7. Interno chiesa con esempio di archi aggettanti (blu) rispetto all'altezza delle volte (arancione).

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., Lerones, 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 Backpacked 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.