Segmentazione semantica neuro-simbolica di nuvole di punti per il patrimonio architettonico mediante Logic Tensor Network

Francesca Matrone[1*] [0000-0002-9160-1674], Shuming Si[2] [0009-0005-2347-5580], Fabio Gigante[2], Francesco Manigrasso[2] [0000-0002-4151-8880], Lia Morra[2] [0000-0003-2122-7178]

[1] Dipartimento di Ingegneria dell'Ambiente, del Territorio e delle Infrastrutture (DIATI), Politecnico di Torino, C.so Duca degli Abruzzi 24, 10129, Torino, francesca.matrone@polito.it
[2] Dipartimento di Automatica e Informatica (DAUIN), Politecnico di Torino, C.so Duca degli Abruzzi 24, 10129, Torino, (shumin.si, francesco.manigrasso, lia.morra)@polito.it

1 Introduzione

La segmentazione semantica delle nuvole di punti rappresenta un passaggio fondamentale per la documentazione digitale del patrimonio architettonico e per lo sviluppo di processi automatizzati scan-to-HBIM. Sebbene le recenti tecniche di Deep Learning abbiano migliorato significativamente il riconoscimento degli elementi architettonici, permangono criticità legate alla complessità geometrica, alla somiglianza tra categorie semantiche e allo sbilanciamento delle classi nei dataset di addestramento. Inoltre, gli approcci esclusivamente neurali apprendono implicitamente le relazioni architettoniche, senza sfruttare direttamente conoscenze geometriche e spaziali esplicite o regole logiche. Il presente lavoro propone un approccio neuro-simbolico che integra Deep Learning e logica differenziabile per migliorare l'accuratezza e la coerenza della segmentazione semantica del patrimonio costruito.

2 Metodologia

L'approccio proposto, denominato PTv3-LTN, combina Point Transformer V3 (PTv3) [1], utilizzato per la classificazione semantica dei punti, con le Logic Tensor Networks (LTN) [2], che consentono di integrare conoscenze architettoniche sotto forma di regole logiche differenziabili durante l'addestramento. Le regole sono definite mediante quattro descrittori geometrici locali (omnivarianza, variazione superficiale, planarità e verticalità) e comprendono vincoli puntuali (*pointwise*), relativi alla compatibilità tra geometria e classi semantiche, e relazionali (*pairwise*), basati sulle proprietà geometriche e spaziali di punti vicini. Per limitare fenomeni di *shortcut learning* sono introdotti meccanismi di regolarizzazione dell'entropia e ricostruzione degli attributi geometrici. L'addestramento prevede due fasi: una prima ottimizzazione con 20 regole candidate e una successiva selezione delle 10 più idonee, utilizzate nella seconda fase. La sperimentazione è condotta sul benchmark ArCH, considerando dieci classi semantiche e due scene di test relative al Sacro Monte di Varallo (SMV) e al Sacro Monte di Ghiffa (SMG).

3 Risultati

Il confronto con la baseline PTv3 evidenzia un miglioramento complessivo delle prestazioni mediante l'integrazione delle LTN (Fig. 1). Considerando congiuntamente le due scene di test, la *mean Intersection over Union* (mIoU) aumenta dal 65,89% al 67,64% (+1,75 punti percentuali), mentre la *mean Accuracy* passa dal 76,28% al 78,35% e l'*Overall Accuracy* dall'85,79% all'87,03%. I miglioramenti della mIoU risultano più marcati nella scena SMV (+1,92 punti) rispetto alla SMG (+0,48 punti). L'analisi per classe mostra incrementi particolarmente significativi per pavimenti (+4,62 punti), archi (+3,52) e categoria other (+2,80). Il confronto visivo evidenzia inoltre una migliore identificazione di modanature, archi e altri elementi architettonici complessi, nonché una maggiore precisione nella delimitazione della linea di demarcazione tra superfici quali muri e volte. Permangono tuttavia difficoltà nella classificazione di alcuni elementi decorativi e porzioni delle coperture. Nel complesso, i risultati confermano le potenzialità dell'approccio neuro-simbolico nell'integrare esplicitamente conoscenze architettoniche nei modelli di Deep Learning, oltre a supportare la classificazione di elementi (annotati nella classe "other") scarsamente rappresentati nelle scene di addestramento, come grondaie e balaustre.

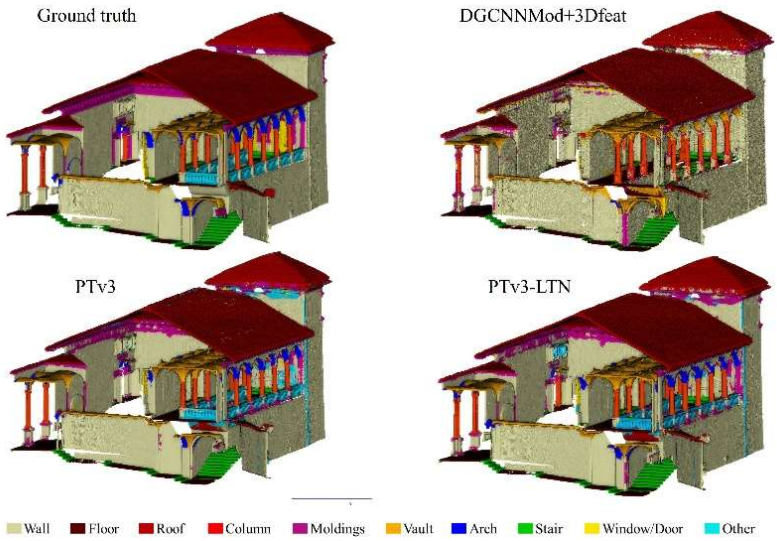

Figura 1. Confronto tra la *ground truth* e le nuvole di punti predette.

Riferimenti bibliografici

1. Wu, X., Jiang, L., Wang, P.-S., Liu, Z., Liu, X., Qiao, Y., Ouyang, W., He, T., & Zhao, H.. Point Transformer V3: Simpler, Faster, Stronger. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 4840–4851 (2024).
2. Badreddine, S., Garcez, A. d'Avila, Serafini, L., & Spranger, M.. Logic Tensor Networks. Artificial Intelligence, 303, 103649 (2022).