

## **Titolo: Tecniche di deep learning per l'analisi automatica di point cloud**

### **Progetto di ricerca e piano di attività**

Nell'ambito della computer vision si sta assistendo all'introduzione sul mercato di sensori ottici 3D caratterizzati da prestazioni sempre più elevate unitamente a costi e ingombri contenuti. Indipendentemente dalla tecnologia di sensing 3D, che può essere, ad esempio, di tipo stereo, TOF, LIDAR o a triangolazione laser, questi sensori forniscono in uscita un insieme non-organizzato di misure tridimensionali (ovvero coordinate  $x,y,z$ ) cui si fa riferimento con il termine di *point-cloud* (nuvola di punti). La natura non organizzata di tali strutture dati, differenti dalle griglie bidimensionali di pixel così efficacemente elaborate dalle CNN, rende non immediata l'applicazione del paradigma del deep learning alle point cloud, sebbene i ricercatori abbiano sin da subito colto le grandi potenzialità degli approcci fortemente data-driven anche nel dominio dell'elaborazione di dati provenienti da sensori ottici 3D. I primi passi della ricerca in questo ambito scientifico, cui si fa talvolta riferimento come 3D Deep Learning, sono stati caratterizzati dal tentativo di normalizzare la rappresentazione grezza dei dati in forme idonee all'utilizzo di architetture neurali standard, quali reti fully-connected e convoluzionali. Ciò tuttavia ha implicato la definizione a priori di tali funzioni di normalizzazione, e quindi una dipartita dal paradigma di training end-to-end, ovvero dai dati grezzi alle uscite, che è stato alla base del successo del deep learning nelle applicazioni di elaborazione delle immagini. Più recentemente, tuttavia, sono stati sviluppati nuovi approcci che paiono dimostrare l'applicabilità di modelli di apprendimento end-to-end anche nel campo dell'elaborazione di nuvole di punti. Inoltre, le problematiche legate alla difficoltà e il costo di ottenere le informazioni di ground-truth necessarie per l'addestramento supervisionato dei modelli neurali sono ancor più rilevanti nel caso dell'elaborazione di point cloud rispetto al caso più comune delle immagini, in ragione sia della intrinseca laboriosità dell'etichettamento di dati 3D sia della sostanziale assenza –allo stato attuale- di tool software specifici. Conseguentemente, nel dominio del 3D Deep Learning appare particolarmente auspicabile lo sviluppo di opportuni tecniche di transfer learning ovvero di paradigmi di apprendimento *unsupervised* o *self-supervised*. In letteratura iniziano quindi ad essere documentati i primi approcci che, da un lato, cercano di realizzare procedure di apprendimento quanto più possibile scevre da scelte *handcrafted* in relazione alla rappresentazione dei dati, dall'altro di definire metodologie di apprendimento senza supervisione nonché di valutare le prestazioni in *transfer learning* (ovvero al variare del dataset di test) dei modelli addestrati. In ragione della giovinezza e della variegata natura degli approcci proposti in letteratura, nel dominio del deep learning applicato all'elaborazione di point-cloud si sta oggi assistendo ai primissimi tentativi di trasferimento tecnologico verso il mondo delle applicazioni industriali. Alla luce delle ben evidenti rilevanti potenzialità applicative del 3D Deep Learning, scopo di questo progetto di ricerca è muovere alcuni passi avanti verso il sopracitato obiettivo di trasferimento tecnologico dal mondo della ricerca a casi di studio industriali caratterizzati da valide opportunità di mercato individuati da Injenia.

Nell'ambito del progetto di ricerca collegato al presente assegno di ricerca si è così stabilito di dedicare la prima parte delle attività ad uno studio approfondito delle più recenti tecniche di 3D Deep Learning, prestando particolare attenzione agli approcci *end-to-end* e basati su paradigmi di apprendimento *unsupervised/self-supervised*, così come al potenziale delle proposte in relazione a scenari di *transfer learning*. L'obiettivo di questa fase di studio è analizzare approfonditamente la più recente letteratura del 3D Deep Learning allo scopo di individuare le proposte che risultano potenzialmente più promettenti in relazione agli scenari applicativi di maggior interesse per Injenia. Questa fase si prevede possa avere una durata di circa 1 mese e produrre come risultato la selezione delle tecniche potenzialmente più interessanti. In questo ambito si evidenzia come particolare attenzione sarà dedicata ad un approccio innovativo proposto dal DISI in una recente pubblicazione<sup>1</sup> presso la conferenza internazionale ICCV 2019. In tale lavoro si propone di usare un nuovissimo framework di *Geometric Deep Learning* (<http://geometricdeeplearning.com/>), ovvero le *Spherical CNN*<sup>2</sup> (reti neurali convoluzionali che lavorano su manifold quali  $S^2$  e  $SO(3)$  invece che su spazi Euclidei) per gestire in maniera *data-driven*, ovvero, diversamente da tutta la letteratura precedente, senza scelte *hand-crafted*, il problema della normalizzazione dell'orientamento delle superfici 3D rispetto al sistema di riferimento solidale con il sensore utilizzato per l'acquisizione della point cloud. Inoltre, per evitare la dipendenza da informazioni di ground-truth difficilmente ottenibili negli scenari applicativi reali, la proposta del DISI adotta un paradigma di apprendimento totalmente *unsupervised*.

Successivamente, l'attività di ricerca sarà focalizzata, per circa 9 mesi, sullo sviluppo di una tecnica (o una famiglia di tecniche) innovative di base che, facendo leva sulla recente letteratura del 3D Deep Learning, siano in grado di affrontare una varietà di task legati all'analisi automatica di point cloud. Esempi di tali task sono il matching di superfici, in scenari di ricostruzione 3D e riconoscimento oggetti, la segmentazione semantica e la ricerca di modelli 3D all'interno di archivi (Shape Retrieval). Sostanzialmente, l'obiettivo dello sviluppo è quello di definire un nuovo framework di 3D Deep Learning che si basi il più possibile su un paradigma di apprendimento *end-to-end* e *unsupervised* nonché supporti il trasferimento della conoscenza fra differenti dataset.

Infine, gli ultimi due mesi di attività saranno dedicati a testare il framework sviluppato in uno o due casi di studio proposti da Injenia.

Come risultati del progetto ci si attende sia lo sviluppo di un framework software di 3D Deep Learning sia la stesura di almeno una pubblicazione scientifica in una prestigiosa sede internazionale.

---

<sup>1</sup> Riccardo Spezialetti, Samuele Salti, Luigi Di Stefano, "Learning an Effective Equivariant 3D Descriptor Without Supervision", ICCV 2019.

<sup>2</sup> Taco S Cohen, Mario Geiger, Jonas Kohler, and Max Welling. "Spherical CNNs", arXiv preprint arXiv:1801.10130, 2018.