

Tecniche di deep learning per la ricostruzione 3D di oggetti non-lambertiani – Progetto di Ricerca e Piano di Attività

Le tecniche di deep learning sviluppate negli ultimi anni hanno portato a risultati davvero molto significativi nell'ambito della ricostruzione della struttura 3D di una scena a partire da immagini. In particolare, il deep learning ha prodotto avanzamenti sostanziali nell'ambito della stima di informazioni di profondità a partire dal classico approccio stereo binoculare così come per via del ben più recente paradigma monoculare (*depth-from-mono*). Inoltre, nel 2020 è stato proposto un approccio fortemente innovativo, detto NEural Radiance Fields (NERF), che, formulando in maniera continua il problema della consistenza fotometrica di più viste di una scena, consente di stimarne la struttura 3D e sintetizzare nuove viste a risoluzione arbitraria.

Nonostante questi avanzamenti, in alcuni aspetti assimilabili a veri e propri *breakthrough* della ricerca in computer vision, rimangono tutt'ora aperte alcune problematiche fondamentali che, sin dagli albori, hanno afflitto le tecniche di ricostruzione 3D da immagini. In particolare, queste tecniche hanno sempre mostrato difficoltà nello stimare la struttura 3D di oggetti caratterizzati da superfici non-lambertiane (cioè non uniformemente riflettenti), quali ad esempio superfici metalliche speculari, oggetti trasparenti e oggetti riflettenti (specchi). Tali problematiche si manifestano sia in caso di utilizzo di algoritmi tradizionali di corrispondenza stereo sia qualora l'approccio perseguito sfrutti deep neural networks (deep stereo). Sono stati recentemente introdotti in letteratura alcuni nuovissimi dataset caratterizzati dalla presenza di immagini, sia di training sia di test, idonee allo studio e sviluppo di una nuova generazione di tecniche per la ricostruzione 3D che, in presenza delle problematiche precedentemente evidenziate, possa fornire risultati decisamente migliori rispetto all'attuale stato dell'arte.

Scopo primario del presente progetto di ricerca è quello di utilizzare i nuovi dataset per la stima di informazioni 3D da immagini recentemente resi disponibili in letteratura allo scopo di sviluppare nuove architetture di deep learning in grado di ottenere risultati significativamente migliorativi rispetto allo stato dell'arte in relazione alle problematiche precedentemente evidenziate. In particolare, in questo ambito di ricerca, saranno sviluppate sia architetture deep per la visione stereo binoculare (*deep stereo*) sia architetture per approcci di tipo "*depth from mono*". Inoltre, saranno sviluppate sia architetture orientate alla generalizzazione a partire da una fase iniziale di *pre-training* (su dati sintetici e/o reali) sia architetture concepite per l'adattamento *online* realizzato mediante tecniche di addestramento che non richiedono *labels* di ground-truth, ovvero tecniche di *self-supervised* learning oppure basate sull'utilizzo di *proxy-labels*. Infine, ulteriore obiettivo del progetto è lo studio e sviluppo di approcci innovativi di tipo *Neural Radiance Field* che possano integrare in maniera coerente informazioni provenienti da più sensori visuali differenti, in maniera da realizzare una ricostruzione della scena più ricca e completa, potenzialmente in grado di evidenziare alcune caratteristiche dei materiali di cui sono costituiti gli oggetti in essa presenti.

Il piano di attività prevede una fase iniziale, della durata di circa un mese, in cui saranno analizzati e classificati i dataset presenti in letteratura potenzialmente idonei allo sviluppo delle architetture neurali che costituiscono l'obiettivo primario del presente progetto. In particolare, saranno

valutate, per gli scenari e i dataset di interesse per il progetto, la disponibilità di immagini di training, sia con ground-truth sia prive di tali informazioni, nonché di immagini di test utilizzabili per la valutazione quantitativa dei risultati. Successivamente, sarà condotta l'attività di sviluppo, portando avanti in parallelo, al fine di sfruttare le potenziali sinergie, sia la ricerca inerente le architetture di *deep stereo* sia quello di tipo *depth-from-mono*. In tale ambito, si ritiene che possa rivelarsi fruttuosa l'intuizione che, in caso di oggetti non-lambertiani, una miglior ricostruzione 3D possa derivare dall'integrazione nei modelli deep di layer in grado di apprendere informazioni di natura semantica circa la natura dei materiali osservati. Si stima di 8 mesi l'attività necessaria allo sviluppo delle nuove architetture "*material-aware*" di tipo *deep stereo* e *depth-from-mono*. Infine, gli ultimi 3 mesi del progetto saranno destinati all'investigazione e possibile sviluppo di un nuovo approccio di tipo NERF basato sull'integrazione di informazioni visuali provenienti da sensori differenti.