

DL Instance Segmentation per applicazioni di bin picking

Vittorio Sala, Andrea Mannella

iIMAGE S s.p.a.

vittorio.sala@imagesspa.it, andrea.mannella@imagesspa.it

Abstract

Oggi giorno, l'utilizzo di reti neurali convolutive per l'elaborazione di immagini sta diventando sempre più comune in applicazioni industriali. In questo studio, si è voluto indagare come il contributo di reti neurali per instance segmentation aiuti i sistemi 3D nell'individuazione di oggetti in applicazioni di bin picking. I sistemi 3D incontrano difficoltà nel riconoscimento di oggetti allineati e adiacenti l'uno all'altro: attraverso le reti di instance segmentation è possibile separare i singoli oggetti (con bounding box) e localizzarli con precisione (attraverso maschere di segmentazione).

1 Introduzione

Questo studio nasce con l'intento di migliorare il sistema di random bin picking risolvendo le criticità di rilevamento di pezzi impaccati, tipici dei sistemi di rilevazione 3D. Nelle Figure 1 e 2, sono mostrati dei blocchetti di legno posizionati all'interno di un cassone e in procinto di essere prelevati. Nell'immagine 2D monocromatica (Figura 1) i singoli blocchetti sono distinguibili grazie agli spigoli e alla differente colorazione. Nell'immagine di profondità (Figura 2), i singoli blocchi non sono identificabili in quanto vengono rinvenuti tutti alla medesima distanza dal sensore, e la zona dei blocchetti è di conseguenza tutta mappata allo stesso modo.

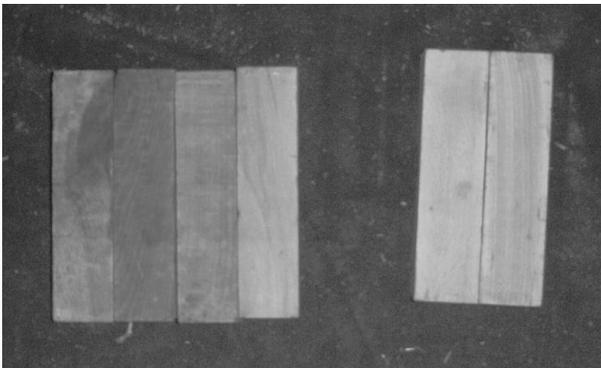


Figura 1. Immagine 2D monocromatica di alcuni blocchi di legno da prelevare attraverso bin picking.

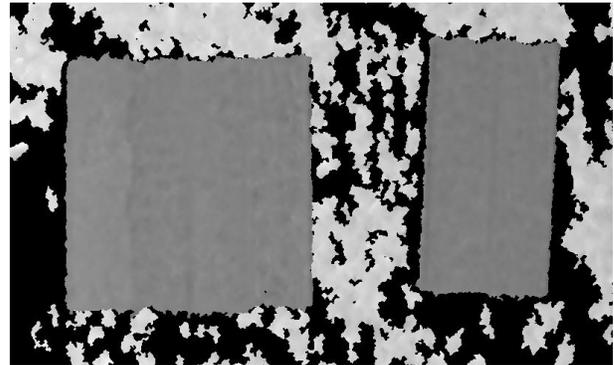


Figura 2. Immagine di profondità dei medesimi blocchetti di legno. Gli spigoli dei singoli blocchetti non sono distinguibili.

Le reti neurali di instance segmentation possono aiutare i sistemi 3D nella localizzazione degli oggetti, elaborando immagini bidimensionali [Hafiz *et al.*, 2020]. Esse individuano i singoli oggetti attraverso bounding box che circondano gli oggetti stessi, e definiscono con precisione la loro localizzazione attraverso maschere di segmentazione [Fu *et al.*, 2019].

2 Preparazione del dataset e addestramento

Al fine di implementare tale soluzione, si è scelto di acquisire un dataset di circa 250 immagini di blocchetti di legno. Sono presenti sia immagini con blocchetti opportunamente disposti (come in Figura 1) sia immagini con blocchetti disposti alla rinfusa in un cassone per simulare un'applicazione reale di bin picking. Le immagini sono state acquisite con un sensore 3D basato su stereoscopia a proiezione di punti, della famiglia Ensenso. Il cassone ha dimensioni 80x60 cm, e la massima distanza misurata dal sensore è circa 120 cm. Per ogni configurazione dei pezzi è stata acquisita sia l'immagine di profondità sia l'immagine monocromatica allineata pixel a pixel con la precedente.

L'addestramento delle reti neurali per instance segmentation prevede che, oltre alle immagini di addestramento, vengano fornite delle maschere di segmentazione (con relative bounding box) per indicare alla rete gli oggetti da individuare e su cui addestrarsi. Le maschere sono state realizzate con lo

scopo di individuare le singole facce non occluse dei blocchetti di legno. Il dataset completo è stato quindi creato utilizzando il software MVTec DLTool mentre l'addestramento della rete di instance segmentation è stato svolto utilizzando HALCON 21.11 Progress, libreria di machine vision e Deep Learning sviluppata da MVTec gmbh. Tra le reti disponibili all'interno della libreria, si è scelto di utilizzare l'instance segmentation basata su ResNet50, come compromesso tra qualità dei risultati e tempi di elaborazione [He et al., 2016]. La libreria consente di operare un fine tuning degli iper-parametri delle reti: gli addestramenti sono stati svolti partendo da un learning rate pari a 0.005 e implementando una strategia di decadimento, utilizzando un momento pari a 0.9 e un batch size pari a 2, scelto per ottimizzare l'uso della memoria della GPU [Jacobs, 1988].

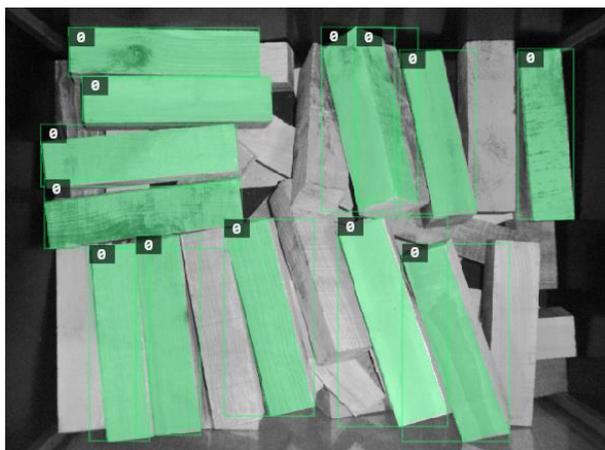


Figure 3. Immagine del dataset fornita per l'addestramento della rete neurale di instance segmentation. In verde, sono disegnate le maschere e le relative bounding box (con assi allineati).

3 Risultati

Considerando un valore di Intersection Over Union (IoU) pari a 0.5, i risultati della rete ottenuti su un set di 38 immagini di test mostrano come la mAP sia pari a 79.96% [Hui, 2019]. All'interno di queste immagini, la rete è stata in grado di identificare correttamente 209 facce dei blocchetti di legno, al contrario però i falsi positivi rinvenuti sono pari a 23 e i falsi negativi ammontano a 45.

Nello studio condotto, i risultati della rete sono stati successivamente filtrati attraverso analisi di forma, dimensione e sovrapposizione delle maschere, con l'ausilio di algoritmi classici di machine vision. Al fine di selezionare i blocchetti di legno più semplici e pratici per il bin picking, si è scelto di escludere le maschere rinvenute in maniera non accurata, non compatibili con la superficie planare della singola faccia. Nella visualizzazione finale, le facce individuate sono rappresentate come contorno della relativa maschera di segmentazione.

I risultati di tutto il processo sono mostrati in Figura 4. Sono riportati alcuni esempi di acquisizioni di blocchetti opportunamente disposti per simulare il problema dei sensori 3D precedentemente citato e da cui nasce questo studio (Figura 4, in alto). Dalle immagini si può notare come la maggior parte dei blocchetti vengano rilevati. La rete è in grado di individuare con maggiore confidenza e precisione i blocchetti più laterali, presumibilmente perché più liberi e con spigoli più identificabili. Anche se nelle immagini non tutti i blocchetti (non occlusi) sono individuati, è possibile comunque selezionarne almeno uno e procedere così, passo dopo passo, allo svuotamento del cassone.

Discorso analogo vale anche per configurazioni randomiche dei campioni, che simulano applicazioni di bin picking (Figura 4, in basso). Le facce non occluse di alcuni blocchetti vengono trovate con una buona precisione, anche se la rete

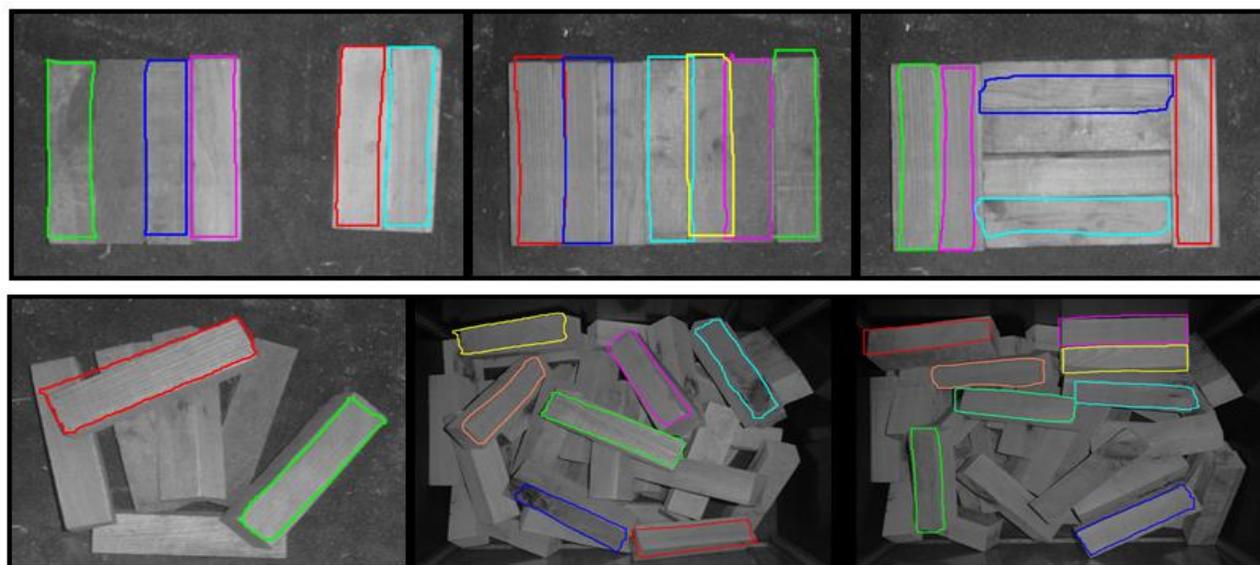


Figura 4. Esempi di risultati su varie disposizioni di blocchetti di legno. I risultati mostrati fanno riferimento all'applicazione della rete neurale di instance segmentation e successivo processing.

non individua tutte le facce disponibili. Il cassone può comunque essere svuotato fintanto che una faccia venga identificata.

4 Conclusioni

Le reti neurali di instance segmentation possono sicuramente aiutare i sistemi 3D nell'individuazione di oggetti, in configurazioni particolari, laddove i sistemi 3D hanno difficoltà. I risultati ottenuti mostrano però che le reti hanno criticità nell'accurata localizzazione degli oggetti: ecco che quindi questa tecnologia si dimostra essere un supporto per i sistemi 3D, ma non una loro alternativa. Inoltre, il dato relativo alla distanza del campione dal sensore 3D è fondamentale per il prelievo del pezzo e in nessun modo può essere ottenuto tramite reti neurali.

Il medesimo studio, in relazione ai campioni a forma di parallelepipedo utilizzati, poteva essere altresì condotto con reti neurali di object detection con assi inclinati, ma si è scelto l'uso della instance segmentation anche come punto di partenza per estendere in futuro la trattazione a campioni asimmetrici, di differenti forme e/o dimensioni. Inoltre, un ulteriore passo successivo è quello di ricorrere a reti neurali di instance segmentation con bounding box orientate, che sembrano garantire migliori prestazioni [Follman e König, 2019].

Riferimenti bibliografici

- [Hafiz *et al.*, 2020] Abdul Mueed Hafiz and Ghulam Mohiuddin Bhat. A survey on instance segmentation: state of the art. *International Journal of Multimedia Information Retrieval*, 9:171–189. July 2020.
- [Fu *et al.*, 2019] Cheng-Yang Fu, Mykhailo Shvets and Alexander C. Berg. RetinaMask: Learning to predict masks improves state-of-the-art single-shot detection for free. In *Computer Vision and Pattern Recognition*, *arXiv:1901.03353*. January 2019.
- [He *et al.*, 2016] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep Residual Learning for Image Recognition. The IEEE Conference on Computer Vision and Pattern Recognition, pp. 770-778, Las Vegas, California, June 2016, IEEE.
- [Jacobs, 1988] Robert A. Jacobs. Increased rates of convergence through learning rate adaptation. *Neural Networks* 1(4): 295-307, 1988.
- [Follman e König, 2019] Patrick Follman and Rebecca König. Oriented Boxes for Accurate Instance Segmentation. In *Computer Vision and Pattern Recognition*, *arXiv:1911.07732*. November 2019.
- [Hui, 2019], Jonathan Hui. mAP (mean Average Precision) for Object Detection. March 2018.