Cognitive Grasping System: un caso applicativo di Reti Neurali Convoluzionali per una manipolazione industriale autonoma e robusta.

Cristina Cristalli, Matteo Mazzanti, Eric Tondelli
Ricerca ed Innovazione, Loccioni
c.cristalli@loccioni.com

Abstract

La manipolazione non supervisionata ha consentito lo sviluppo di una vasta gamma di operazioni da far fare ai robot. Tuttavia, percepire l'ambiente esterno è ancora un problema di difficile soluzione nel campo della ricerca della robotica intelligente a causa della mancanza di disponibilità dei modelli degli oggetti, di ambienti non strutturati, e di calcoli che richiedono ancora tempi troppo lunghi. L'intelligenza artificiale utilizzata in questo ambito sta dando buoni risultati e molti sforzi si stanno facendo per adattarla alle esigenze industriali. Questo lavoro vuole essere una testimonianza di quali esigenze ci sono da soddisfare per avere una rapida e positiva applicazione dell'Intelligenza Artificiale in ambienti di produzione nel caso specifico della manipolazione avanzata.

1 Introduzione

In un contesto manifatturiero caratterizzato da un'elevata flessibilità e necessità di produrre prodotti diversi in una stessa linea di produzione, c'è una sempre maggiore richiesta di sistemi automatici altrettanto flessibili ed in grado di auto adattarsi alle esigenze produttive. La crescente diffusione di sistemi di visione artificiale sempre più performanti, di robotica collaborativa sempre più vicina all'operatore e la disponibilità di sistemi embedded e PC con alte prestazioni computazionali, fa sì che l'Intelligenza Artificiale stia passando dai laboratori di ricerca alle linee di produzione al fine di supportare questa trasformazione. Uno dei compiti più difficili da risolvere è quello della manipolazione fatta con robot in modo autonomo [Pratichizzo *et al.*, 2008] e quindi non supervisionato. La percezione del modo con cui afferrare un oggetto fatta in modo automatico, ovvero senza avere informazioni a priori sulle sue dimensioni e orientamento, è infatti un argomento di studio di molti ricercatori. Il lavoro presentato in questo articolo, si basa sugli sviluppi fatti da [Mahler *et al.*, 2017] al fine di utilizzare le "Grasp Quality Convolutional Neural Network" (GQ-CNN) per afferrare oggetti in modo autonomo con robot. Il lavoro svolto ha l'obiettivo di rielaborare quanto reso disponibile dai ricercatori, al fine di renderlo utilizzabile a livello industriale. Questo consiste nel rielaborare gli algoritmi sviluppati ed inserirli in un contesto industriale che va ad ottimizzare le performance del modello inserendolo in un sistema completo composto da telecamera 3D, Robot, pinza e PC. Gli elementi citati sono alla base di un sistema automatico di manipolazione, chiamato Cognitive Grasping System (CGS), capace di manipolare diversi oggetti senza avere informazioni a priori di quel componente, quali ad esempio i disegni CAD.
L'applicazione sviluppata tratta della manipolazione di componenti del settore automobilistico che vanno mossi da una stazione all'altra per fare dei test per controllo qualità. Il Cognitive Grasping System sviluppato è in grado di adattarsi a diversi componenti da manipolare senza la necessità di dover cambiare le coordinate di presa ogni qual volta c'è un nuovo prodotto da manipolare e quindi di ridurre drasticamente l'intervento umano.

2 Overview del Sistema

Il CGS (Fig. 1) è composto principalmente da tre moduli:

1- Il modulo di Intelligenza Artificiale
2- Telecamera 3D
3- Robot e pinza

Figura 1: Sistema di manipolazione non supervisionato.

2.1 Modulo di Intelligenza Artificiale

Il modulo d'intelligenza artificiale è composto principalmente da una rete convoluzionale (CNN). Questa rete, che è stata addestrata su un numero elevato d'immagini di oggetti,

è in grado di capire come generalizzare la presa su oggetti mai visti in precedenza. Il dataset è composto di 1500 modelli di oggetti. In particolare, da 1358 oggetti 3D sintetici, cioè generati direttamente da CAD e 142 Point Cloud acquisite da una telecamera 3D. Questo ci ha permesso di aumentare le caratteristiche di generalizzazione della rete.
A ogni modello sono associate circa 200 possibili pose parallele di presa. Le pose parallele di presa sono ottenute attraverso un algoritmo deterministico. L'algoritmo tiene conto dei vincoli impostati come larghezza massima e dimensione della pinza. In questo modo è possibile ottenere delle pose candidate di presa in funzione della pinza scelta.
La rete è composta da 4 layer convoluzionali utilizzando ReLU come funzione di attivazione, a cui seguono 3 fully connected layer. Inoltre è stato aggiunto un ulterirore fully connected layer z, concatenato con il primo fully-connected layer, che rappresenta la distanza della pinza dalla telecamera. Dal Dataset ottenuto è possibile addestrare la rete convoluzionale che è in grado di trovare la presa parallela migliore. L'input alla rete convoluzionale è rappresentato da un'immagine "depth", acquisita attraverso la telecamera.
L'output della rete è rappresentato da:

1. Una Posa
2. Un indice di qualità

La posa parallela rappresenta la posizione nello spazio di presa dell'oggetto. L'indice di qualità rappresenta un valore di bontà della posa.

2.2 Telecamera 3D

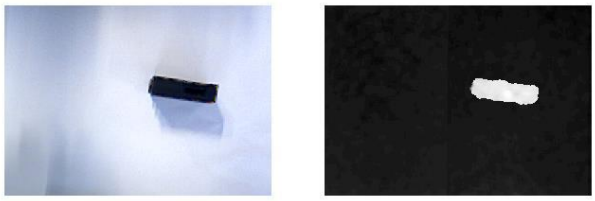
La telecamera 3D è capace di fornire come output una Point Cloud. Questi dati sono poi rielaborati per ottenere un'immagine "depth" (Fig.2).
La telecamera utilizzata è la Ensenso-N35. Questo tipo di telecamera utilizza la tecnica di stereovisione con proiezione di pattern per ottenere la Point Cloud.
La telecamera è computazionalmente pesante se usata direttamente sulla CPU. Per ottenere performance migliori si è quindi utilizzata l'accelerazione della GPU.
Per standardizzare e normalizzare le acquisizioni della telecamera sono stati utilizzati algoritmi di machine vision per migliorare l'immagine acquisita.

2.3 Robot

Diversi tipi di robot sono stati interfacciati nel CGS, quali l'Universal Robot, l'IIWA della Kuka e lo Yumi della ABB.
I vari robot sono stati equipaggiati con diverse pinze ed il parametro utilizzato dalla CNN è esclusivamente la corsa massima della pinza. Una volta che la rete ha calcolato la migliore posa (Fig.3), vengono applicate delle trasformazioni per ottenere le coordinate di presa da passare al robot.

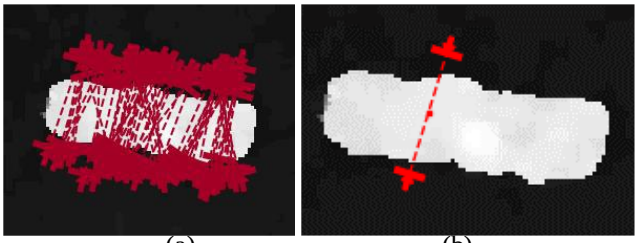
3 Risultati

Il sistema è stato testato utilizzando ROS (Robot Operating System) su un computer con Ubuntu 16.04 con processore Intel i7 Quad-Core e scheda Nvidia Quadro. La rete convoluzionale è stata realizzata utilizzando Tensorflow e Python.
Il tempo computazionale della rete convoluzionale è di circa 2s e l'accuratezza raggiunta dalla rete effettuando delle prove in laboratorio con oggetti diversi e di circa il 97%.
I motivi principali di errore da parte della rete convoluzionale sono dovuti ad un'acquisizione non corretta dell'immagine da parte della telecamera. Ad esempio oggetti troppo sottili non sono acquisiti in modo corretto, causando un errato calcolo della posa di presa.



(a) (b)

Figura 2: Immagine 2D dell'oggetto da manipolare (a) e immagine "depth" (b)



(a) (b)

Figura 3: Risultati 1° (a) e 3° (b) iterazione della rete CNN.

4 Conclusioni e prossimi passi

I risultati ottenuti hanno messo in evidenza che la rete CNN addestrata con un database generale è in grado di afferrare diversi oggetti, senza una ulteriore personalizzazione del modello. Gli sforzi fatti per portare il modello in un ambiente industriale hanno interessato l'ottimizzazione dell'algoritmo e delle diverse fasi di preparazione dell'immagine in ingresso e l'interfacciamento con i diversi componenti del sistema. Particolare attenzione è stata rivolta all'Hardware dove far girare il modello CNN e il preprocessamento delle immagini. Nelle prime sperimentazioni si è preferito utilizzare PC con GPU, ma gli sviluppi sono rivolti verso l'uso di sistemi embedded. Questi sviluppi hanno l'obiettivo di diminuire il tempo ciclo dell'operazione e di portare il CGS a diventare un modulo indipendente e flessibile di manipolazione, e un componente essenziale dei Cyber Physical System.

Riferimenti bibliografici

[Pratichizzo *et al.*, 2008] Prattichizzo D., Trinkle J., Grasping. In Springer handbook of robotics, pages 671–700. Springer, 2008.

[Mahler *et al.*, 2017] J. Mahler, J. Liang, S. Niyaz, M. Laskey, R. Doan, X. Liu, J. A. Ojea, and K. Goldberg, "Dex-net 2.0: Deep learning to plan robust grasps with synthetic point clouds and analytic grasp metrics," in Proc. Robotics: Science and Systems (RSS), 2017.