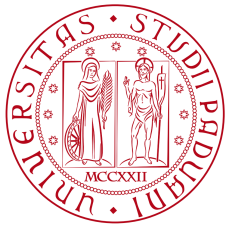


UNIVERSITÀ DEGLI STUDI DI PADOVA

Università degli studi di Padova
Laurea Triennale in Ingegneria Informatica

DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

FUSIONE DI DATI STEREO E TIME-OF-FLIGHT MEDIANTE TECNICHE DI DEEP LEARNING

Relatore: Prof. Pietro Zanuttigh
Correlatore: Ing. Gianluca Agresti

Laureando: Francesco Pham

Anno Accademico 2018-2019
25 settembre 2019

Introduzione

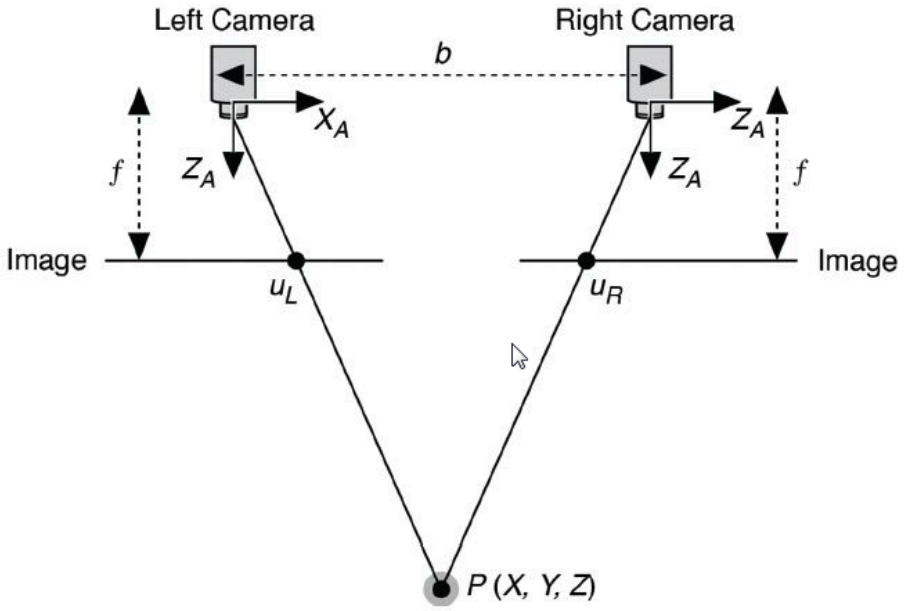
- La stima della profondità di scene tridimensionali rappresenta un problema di forte interesse in molti ambiti, ad esempio:
  - Robotica e automazione
  - Intrattenimento
  - Arte e architettura
- Nel corso degli anni, dispositivi dai costi più ridotti sono stati introdotti nel mercato. Due tipi di sensori in particolare:
  - Il sistema stereo
  - I dispositivi Time-of-Flight

(a) Real-image

(b) Depth-map

Il sistema stereo

- Il sistema di visione stereo consiste nell'acquisire due immagini da una coppia di telecamere che inquadrano la stessa scena.
- Lo stesso punto P viene proiettato nel piano dell'immagine di ciascuna telecamera. I punti risultanti sono chiamati *omologhi*.
- La profondità viene calcolata per triangolazione.
- Principale svantaggio: difficoltà nell'analisi di scene con pattern uniformi o ripetitivi.



Il sensore Time-Of-Flight

Il principio di questa tecnologia è semplice: viene misurato il tempo che un impulso luminoso impiega per viaggiare da una sorgente luminosa ad un oggetto e ritornare al sensore. Limitazioni:

- Misure poco accurate per superfici poco riflettenti o di colore scuro
- Limitata risoluzione spaziale
- Il «multipath error» provocato dalla riflessione multipla del segnale prima di raggiungere il sensore

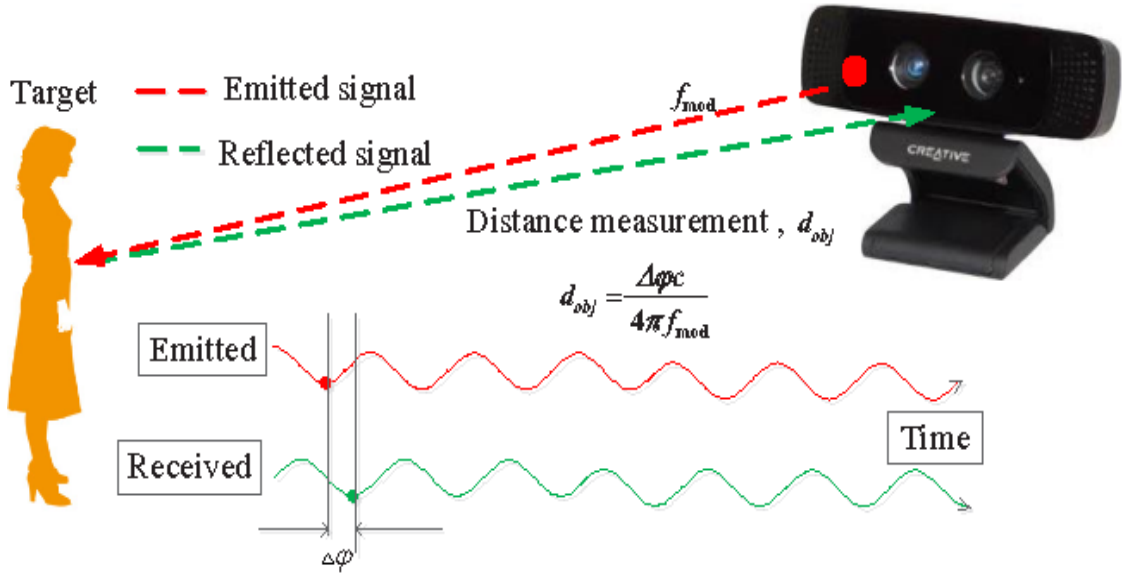
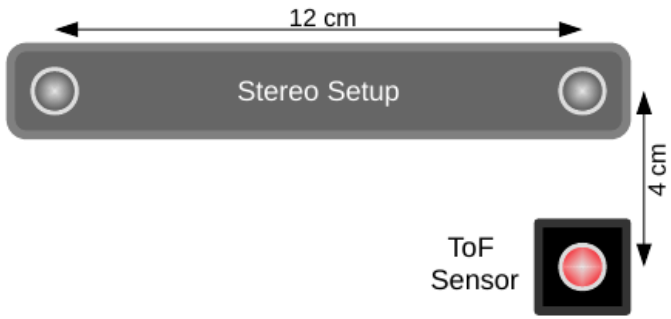
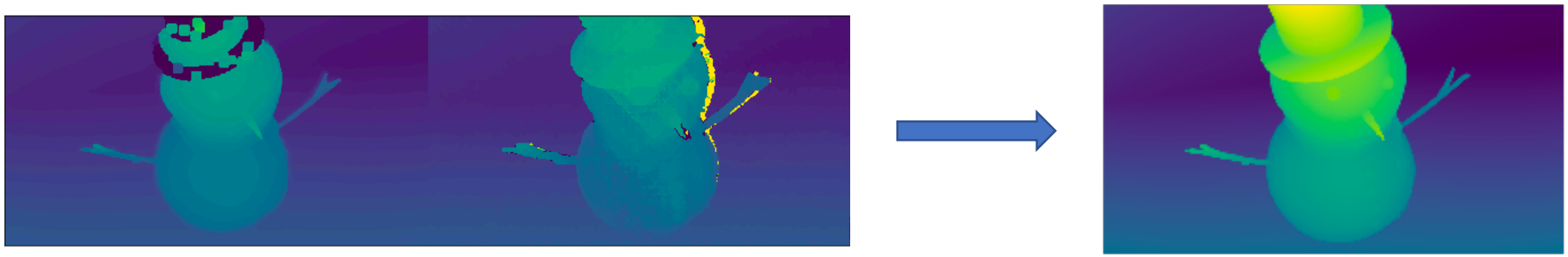


Figure 1. Working principle of ToF ranging camera

Fusione mediante deep learning

- Nel campo della computer vision è stato possibile ottenere progressi notevoli negli ultimi anni grazie al deep learning. In particolare, le reti neurali convoluzionali (CNN) costituiscono lo stato dell'arte nella risoluzione di problemi di visione artificiale.
- L'obiettivo di questa tesi è creare un modello di CNN in grado di fondere i dati acquisiti dal sistema stereo e dal sensore ToF, realizzando una ricostruzione più accurata.





ToF **Stereo**

Il dataset

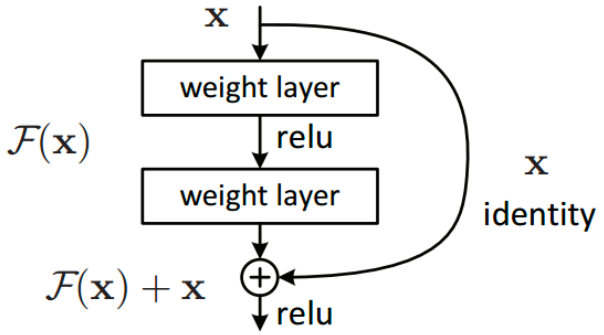
Il dataset consiste in 55 scene 3D differenti simulate con *Blender.*

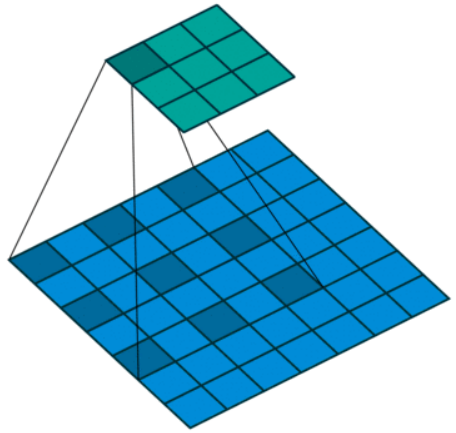
Per poter compiere la fusione è necessario preparare i dati:

Il sensore ToF ha risoluzione nettamente inferiore rispetto allo stereo	→ Interpolazione
Per fare la fusione è necessario che i dati forniti siano nello stesso sistema di riferimento	→ Riproiezione
I due sensori rappresentano i dati in modo differente	→ Calcolo della disparità
Il dataset è limitato. È necessario ampliare il training set per una maggiore robustezza.	→ Data augmentation

Architettura della CNN selezionata 1/2

- Per questo lavoro è stato deciso di provare l'utilizzo di una **rete neurale residuale**. Viene sfruttato il concetto delle *skip connection,* che permette l'apprendimento del contributo incrementale a quanto già appreso negli strati precedenti.
- È stato inoltre deciso di includere nella rete alcuni strati di **convoluzione dilatata**. Permette di catturare più informazioni dall'input espandendo il campo recettivo, pur mantenendo limitato il numero di parametri.





Architettura della CNN selezionata 2/2

Diagramma del blocco residuale

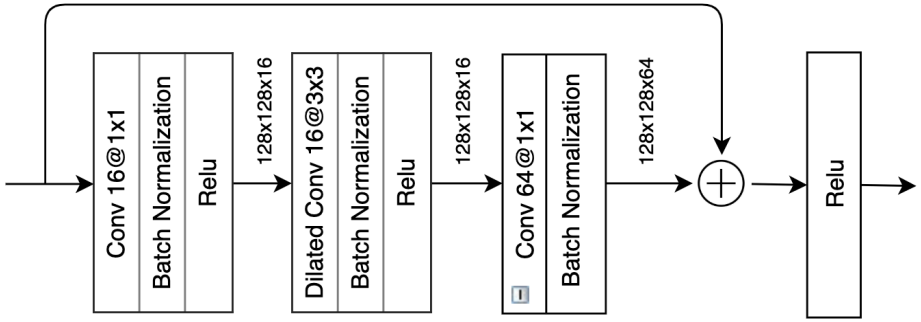
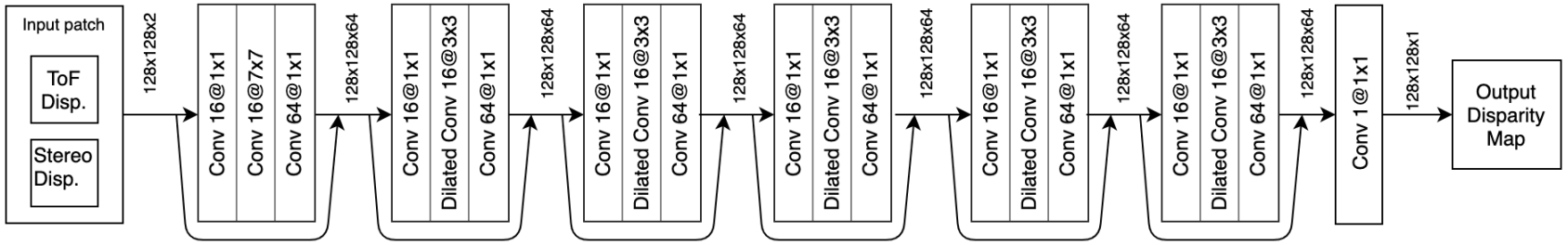
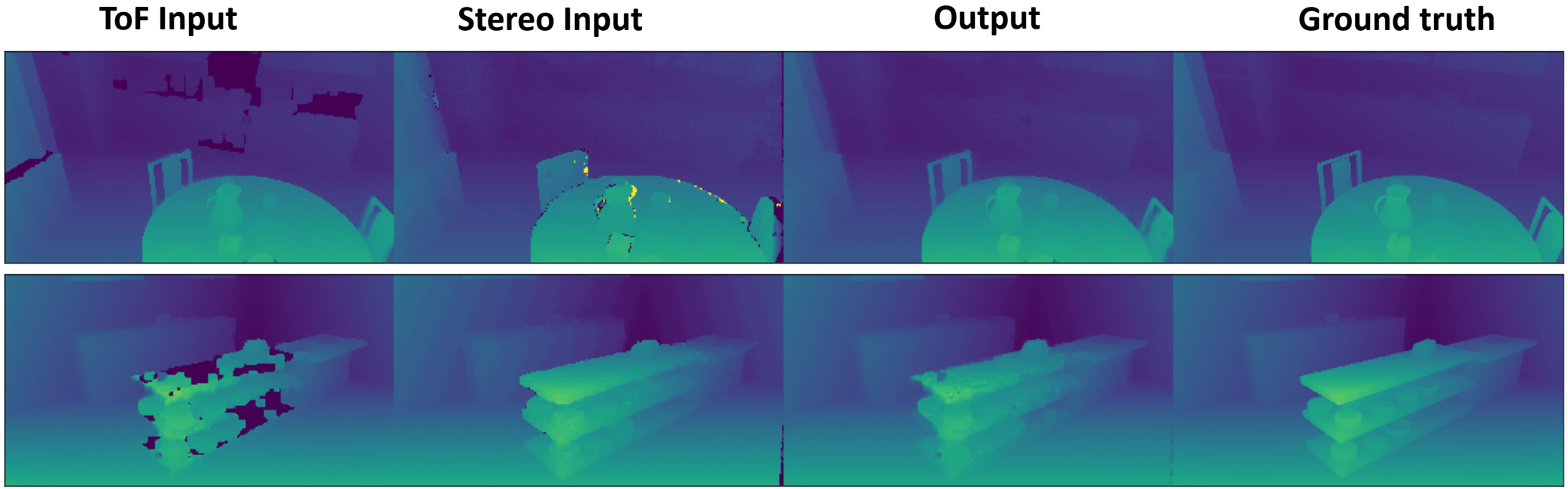


Diagramma completo della rete



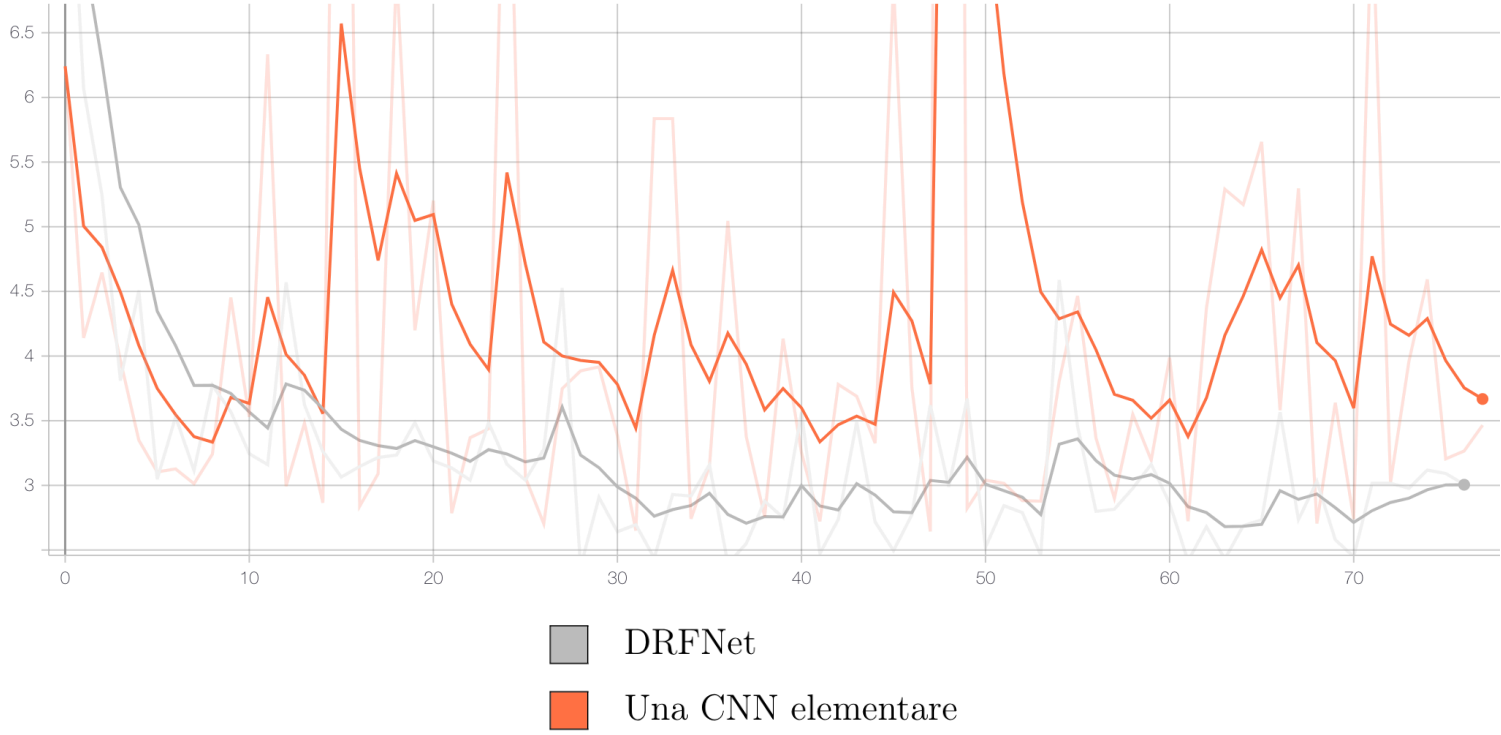
Risultati sperimentali 1/2

- Durata del training: 75 epoche
- Ottimizzatore Adam, learning rate 0.001
- Loss function MSE



Risultati sperimentali 2/2

Proviamo a fare un confronto per vedere gli effetti dei blocchi residuali e delle convoluzioni dilatate sulle performance della CNN



Conclusioni

- Questo lavoro dimostra come il deep learning permetta di realizzare un modello in grado di sfruttare al meglio le informazioni fornite dai due sensori
- Il sistema realizza una ricostruzione più accurata delle strutture tridimensionali della scena catturata.
- Inoltre si è visto come l'utilizzo delle reti neurali residuali assieme alle convoluzioni dilatate abbia apportato benefici sulle performance della rete nella stima della disparità.

Grazie per l'attenzione

Francesco Pham