



UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Università degli studi di Padova
Laurea Triennale in Ingegneria Informatica



DIPARTIMENTO
DI INGEGNERIA
DELL'INFORMAZIONE

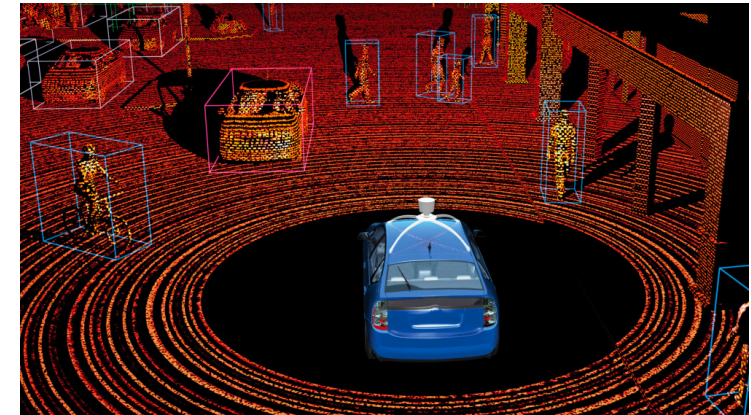
FUSIONE DI DATI STEREO E TIME-OF-FLIGHT MEDIANTE TECNICHE DI DEEP LEARNING

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

Laureando: Francesco Pham

Anno Accademico 2018-2019
25 settembre 2019

- La stima della profondità rappresenta un problema di forte interesse in molti ambiti:
 - Robotica e automazione
 - Intrattenimento
 - Arte e architettura
- Dispositivi dai costi più ridotti sono stati introdotti nel mercato.
 - Il sistema stereo
 - I dispositivi Time-of-Flight



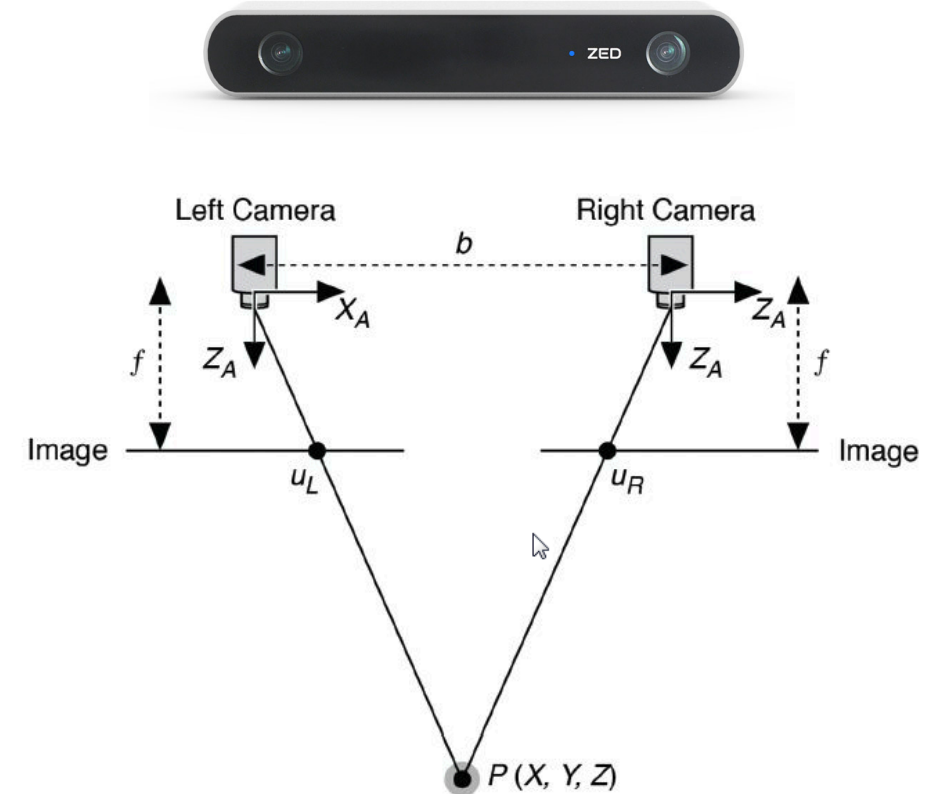
(a) Real-image



(b) Depth-map

Il sistema stereo

- Vengono acquisite 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 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

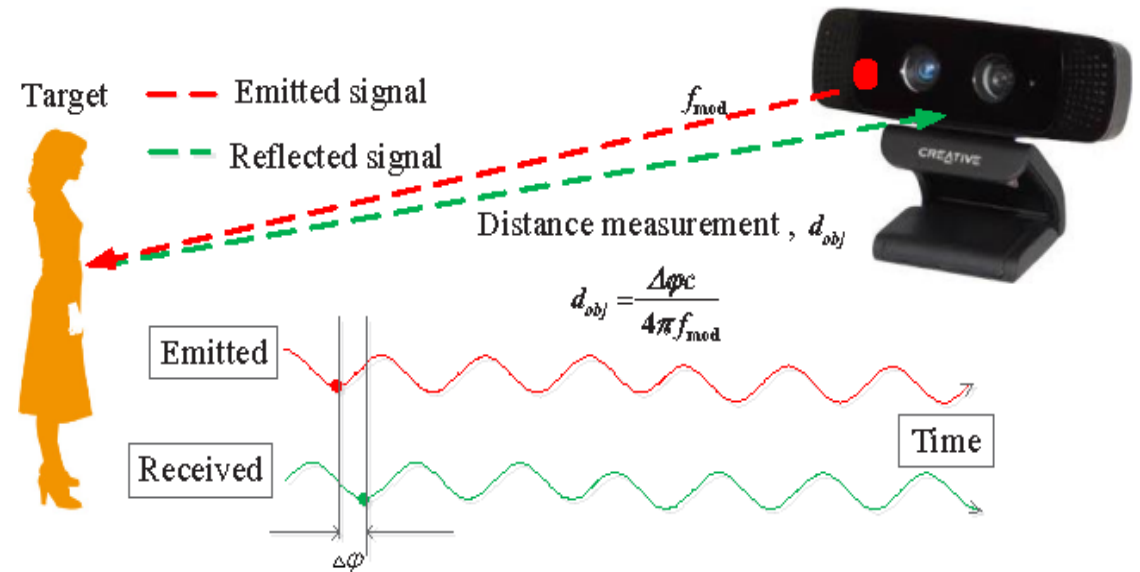


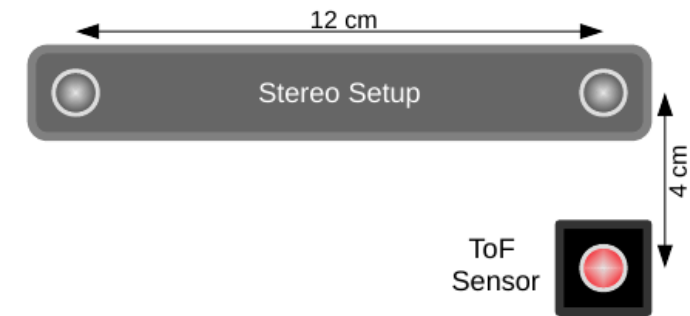
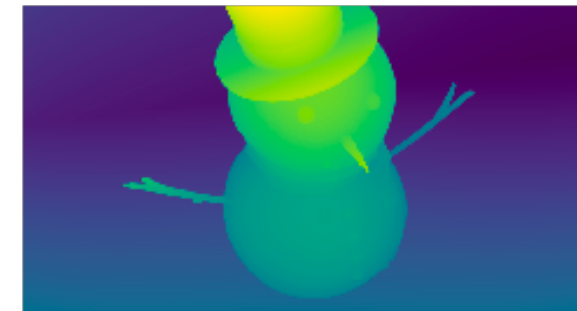
Figure 1. Working principle of ToF ranging camera

- Nel campo della computer vision è stato possibile ottenere progressi notevoli negli ultimi anni grazie al deep learning.
- L'obiettivo di questa tesi è creare un modello di rete neurale convoluzionale 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 preprocessare 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

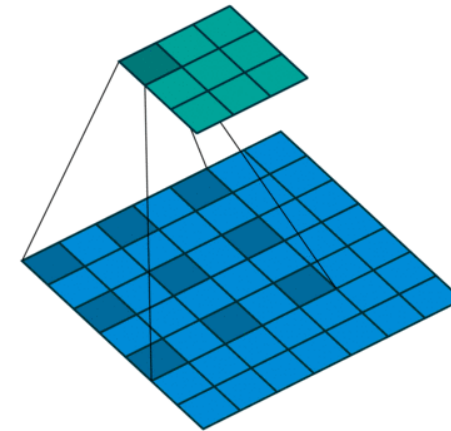
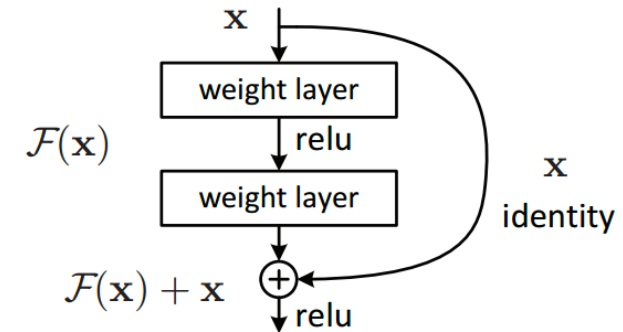
Il dataset è limitato



Data augmentation

Architettura della CNN selezionata 1/2

- È stato deciso di utilizzare 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.
- Sono stati inclusi alcuni strati di **convoluzione dilatata**. La dilatazione permette di catturare più informazioni dall'input espandendo il campo recettivo.



Architettura della CNN selezionata 2/2

Diagramma del blocco residuale

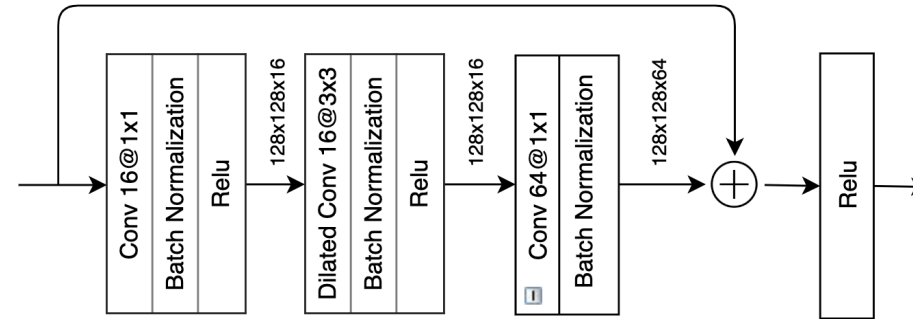
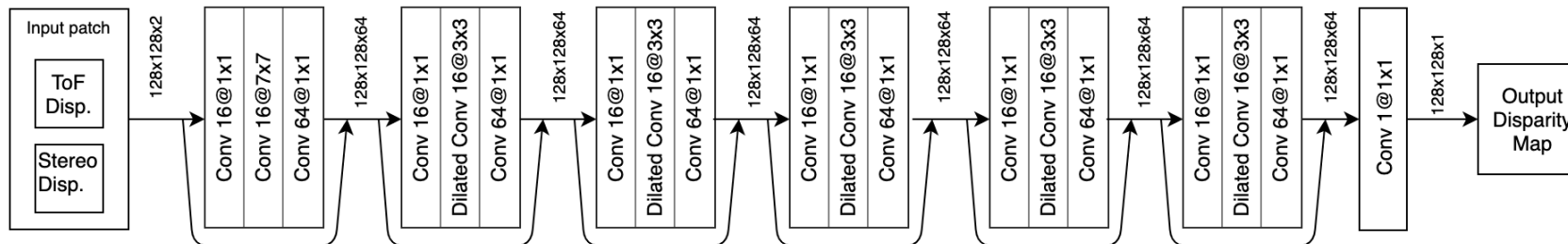


Diagramma completo della rete



Risultati sperimentali 1/2

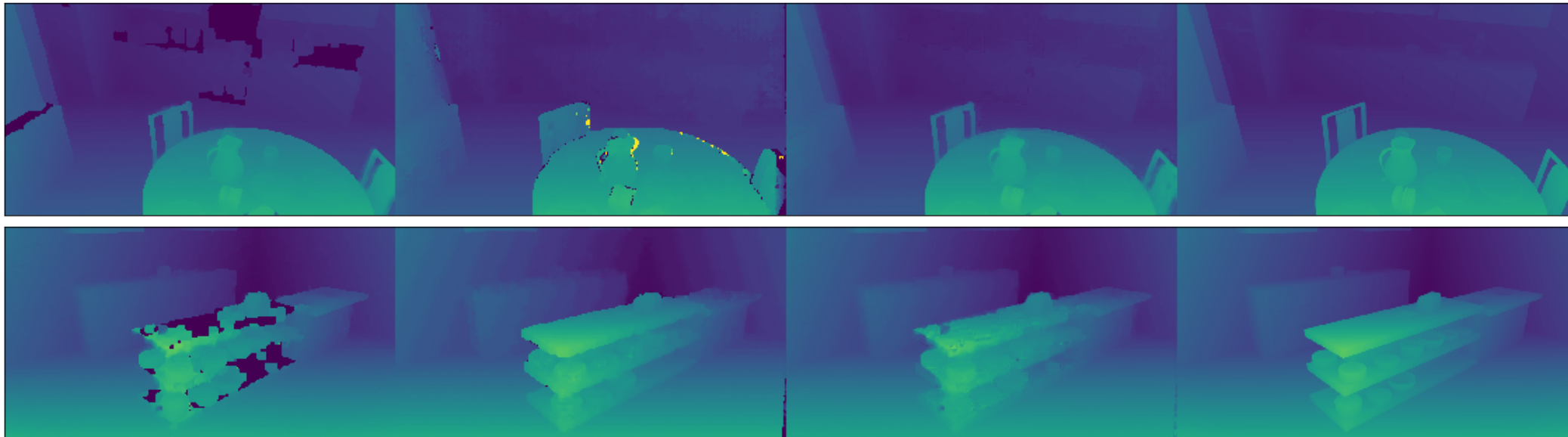
- Durata del training: 75 epoche
- Ottimizzatore Adam, learning rate 0.001
- Loss function MSE

ToF Input

Stereo Input

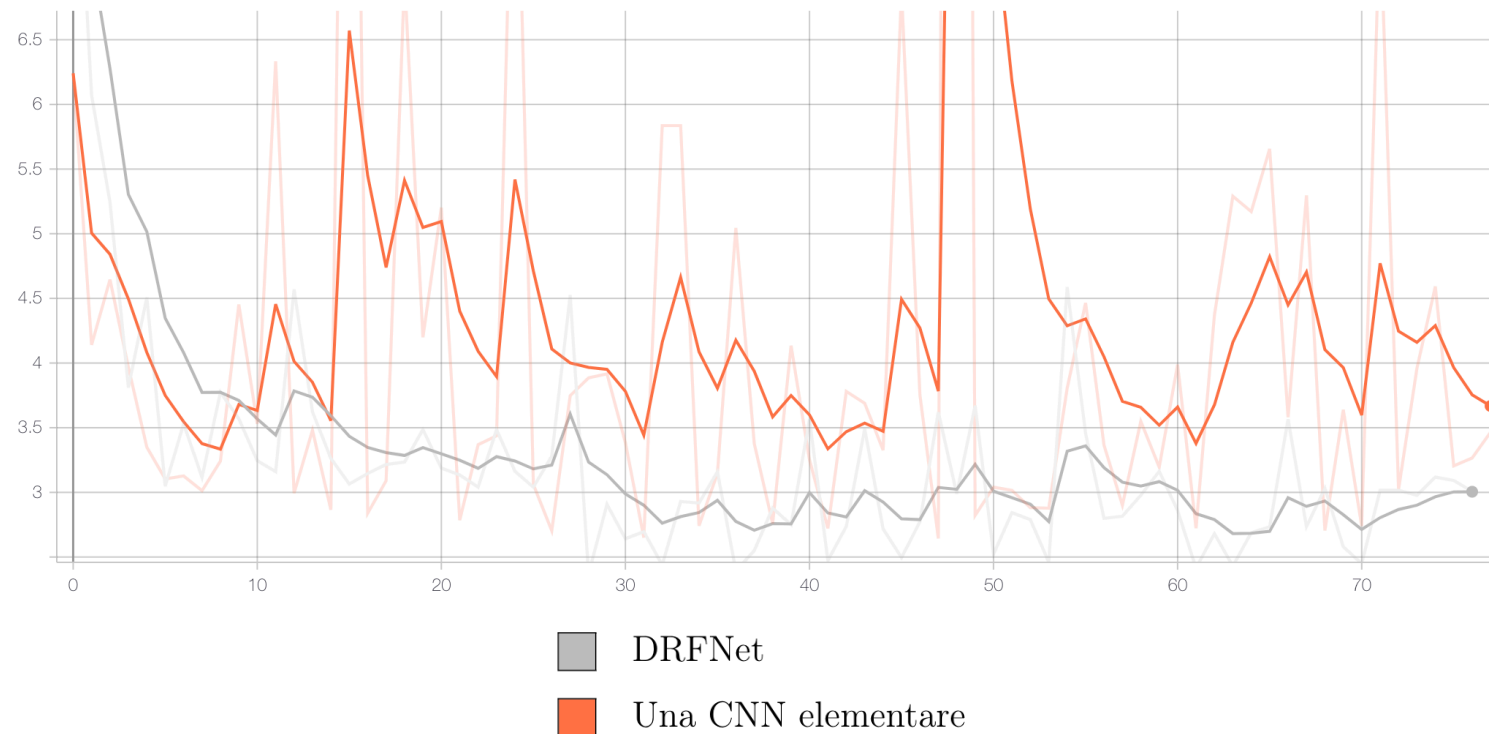
Output

Ground truth



Risultati sperimentali 2/2

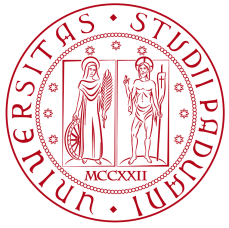
Nella figura seguente viene mostrato un confronto per vedere gli effetti dei blocchi residuali e delle convoluzioni dilatate sulle performance della CNN





Conclusioni

-
- Il deep learning ha permesso 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.
 - L'utilizzo delle reti neurali residuali assieme alle convoluzioni dilatate ha apportato benefici sulle performance della rete.



UNIVERSITÀ
DEGLI STUDI
DI PADOVA



Grazie per l'attenzione

Francesco Pham