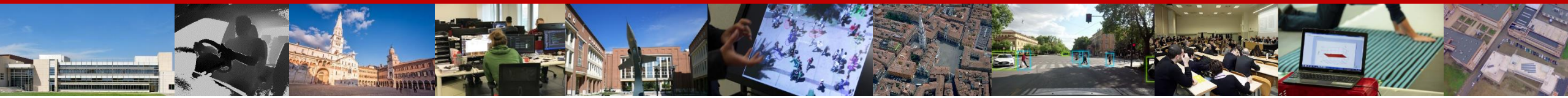
RedVision lab

AImageLAB - Unimore

Riconoscimento di gesti per l'interazione uomo-veicolo mediante sensori infrarossi e 3D



Candidato: Manganaro Fabio

Laurea Magistrale in Ingegneria Informatica

University of Modena and Reggio Emilia, Italy

Relatore: Prof. Roberto Vezzani

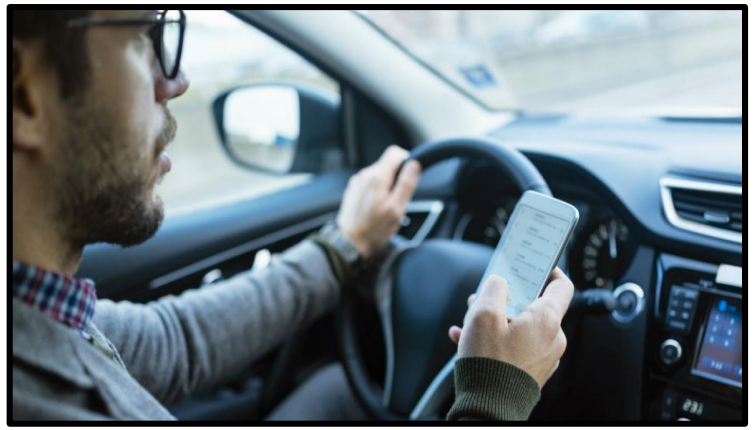
Correlatori: Dott. Guido Borghi
Dott. Stefano Pini

Monitoraggio Stato Guidatore

Più del **90%** degli **incidenti stradali** sono dovuti a **errori umani.**[1]

Una delle **cause principali** è la **Distrazione** del **Driver**:

- Visiva
- Cognitiva
- **Manuale**



1. *EuroNCAP 2025 Roadmap*

Possibile Soluzione

Interazione Uomo-Macchina (HMI) tramite **Natural User Interface** (NUI):

- **No** Dispositivi **Fisici** (manopole, pulsanti, touchscreen)
- Interazione **Intuitiva**

Gesture Recognition

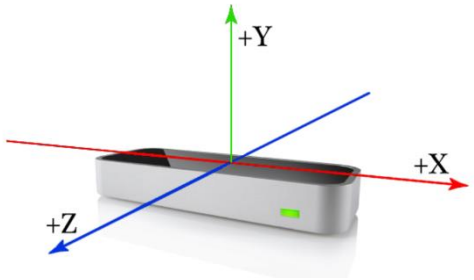
riconoscimento automotatico dei **gesti** mediante sistemi di **visione**

Prototipo Hardware - Scelte Progettuali

Vincoli:

- Posizione e ingombro fisico
- Invarianza Luminosità
- Prestazioni Real Time

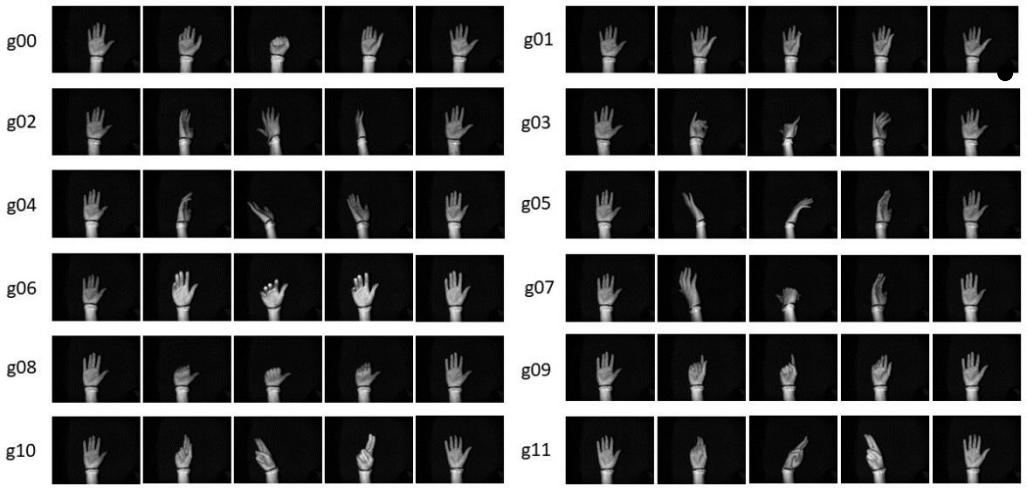
Pico Flexx
ToF



Leap Motion
Stereo-camera infrarossi

RGB cam

Ideazione ed Acquisizione Dataset



12 Gesti Dinamici:

- Fist – g00
- Pinch – g01
- Flip over – g02
- Telephone – g03
- Right Swipe – g04
- Left Swipe – g05
- Top-down swipe – g06
- Bottom-up – g07
- Thumb – g08
- Index – g09
- Clockwise rotation – g10
- Counterclokwise rotation – g11

40 Soggetti

Dati Acquisiti:
IR – RGB – Depth Map – 3D Feature

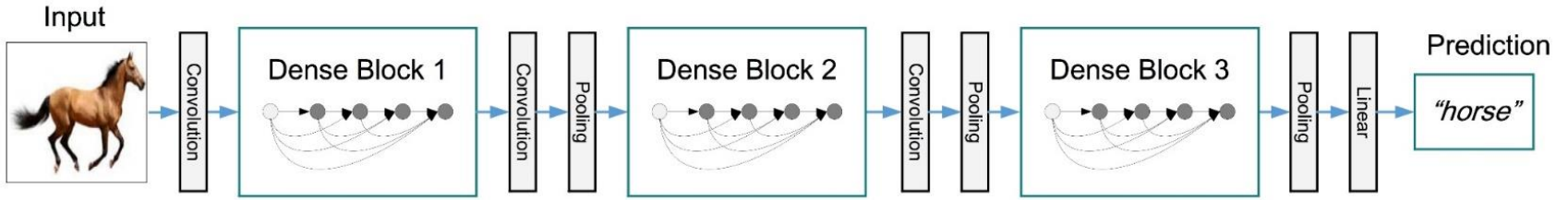
Temporalità gesti

L'**evoluzione temporale** dei gesti dinamici è affrontata lato software **da 3 Architetture**:

1. DenseNet con **Stack di frame**
2. C3D con **convoluzioni 3D** (dimensione tempo)
3. LSTM- **Rete Neurale Ricorrente**

Metodo – 1: CNN

DenseNet[1]



Input:

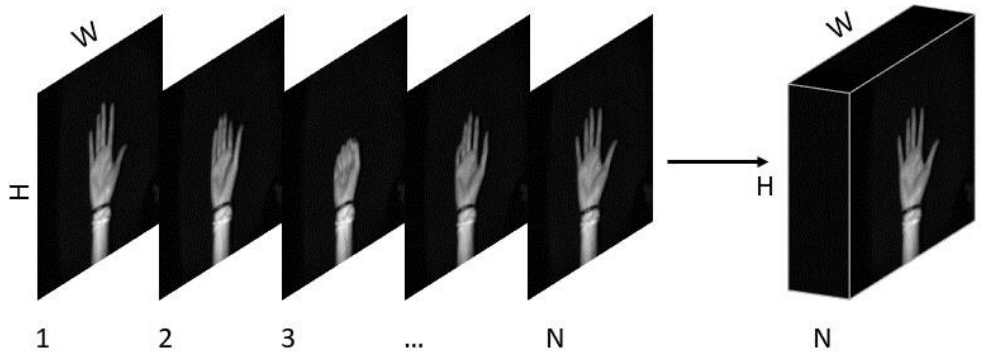
- Depth Map
- Immagini IR
- Immagini RGB/Grayscale

Output:

- Classe Gesture

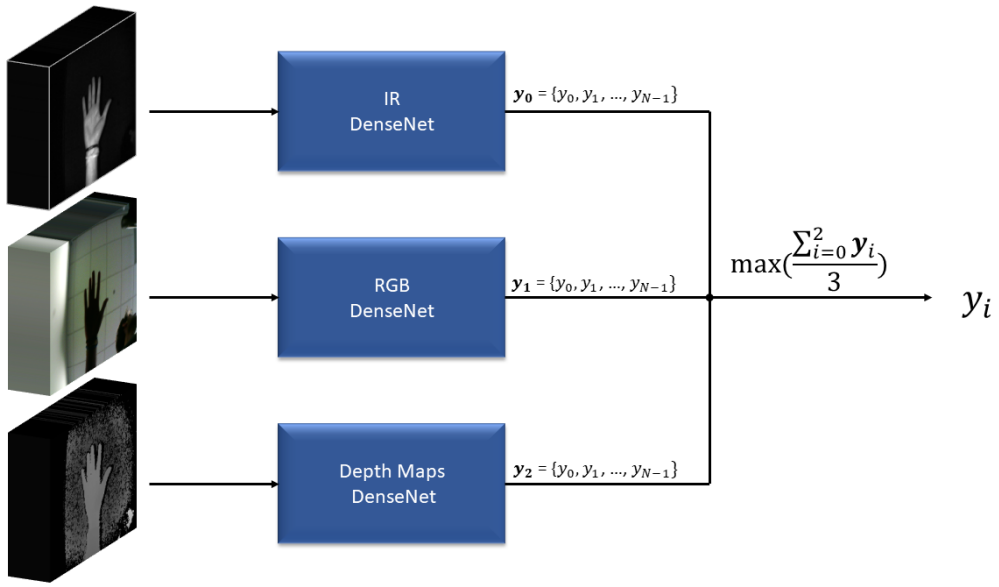
Temporalità:

- Stack



1. Huang, Gao, et al. "Densely connected convolutional networks." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.

Metodo – 2: Architettura Multimodale



Input:

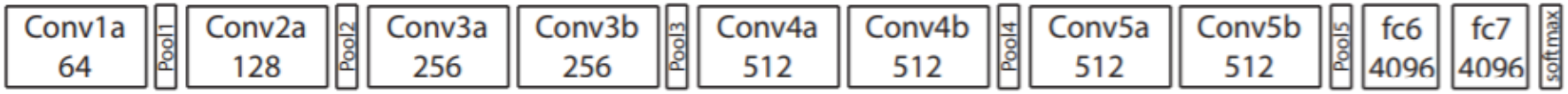
- Depth Map – IR
- Depth Map – RGB/Grayscale
- IR – RGB/Grayscale
- Depth Map – IR – RGB/Grayscale

Output:

- Classe Gesture

Metodo – 3: Reti Neurali convolutive 3D

C3D[1]



Input:

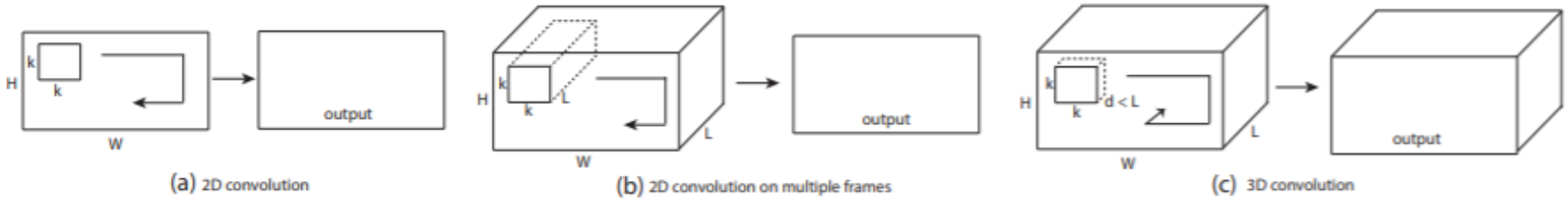
- Depth Map
- Immagini IR
- Immagini RGB/Grayscale

Output:

- Classe Gesture

Temporalità:

- Convoluzioni 3D



1. *Tran, Du, et al. "Learning spatiotemporal features with 3d convolutional networks." Proceedings of the IEEE international conference on computer vision. 2015.*

Metodo – 4: Reti Neurali Ricorrenti

Input:

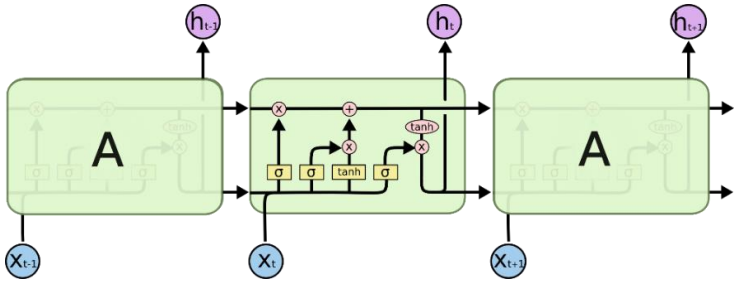
- 3D Feature:
  - Coordinate 3D dei giunti
  - Velocità 3D
  - Imbardata, beccheggio, rollio

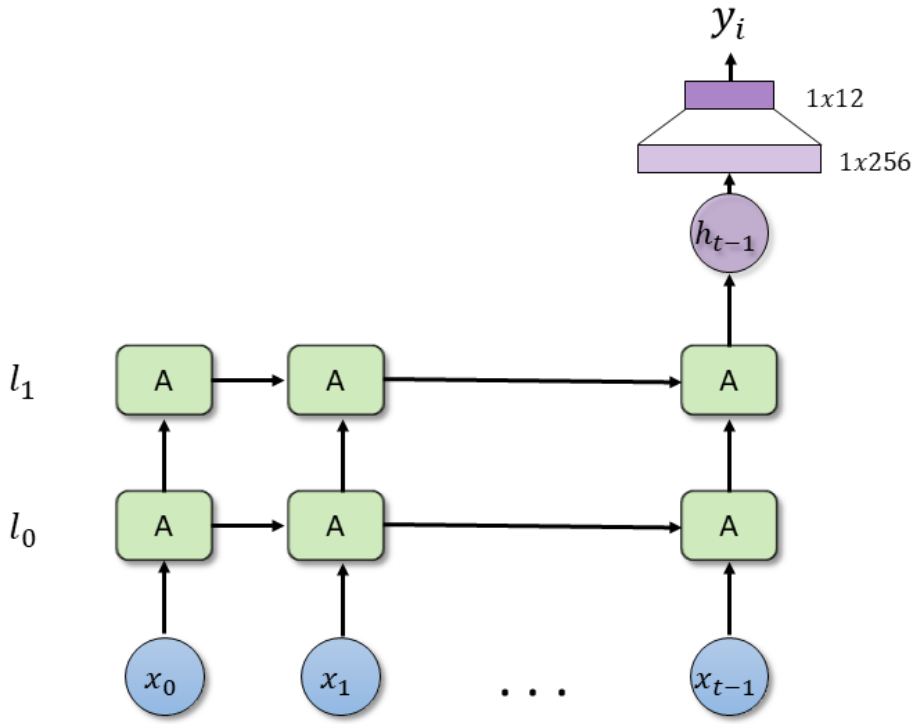
Output:

- Classe Gesture

Temporalità:

- Rete Ricorrenti





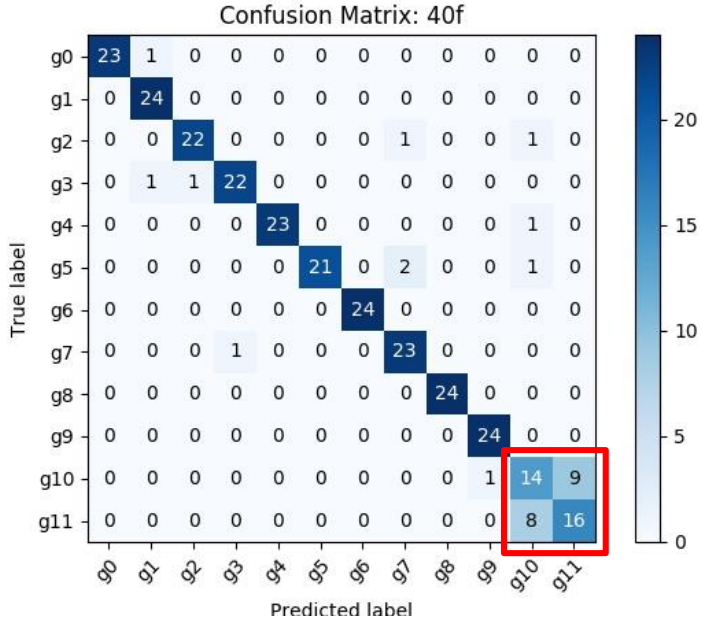
LSTM[1]

1. Gers, Felix A., Jürgen Schmidhuber, and Fred Cummins. "Learning to forget: Continual prediction with LSTM." (1999): 850-855.
2. 12th Eurographics Workshop on 3D Object Retrieval (3DOR), SHREC 2019 Track: online gesture recognition.

Risultati Sperimentali - 1

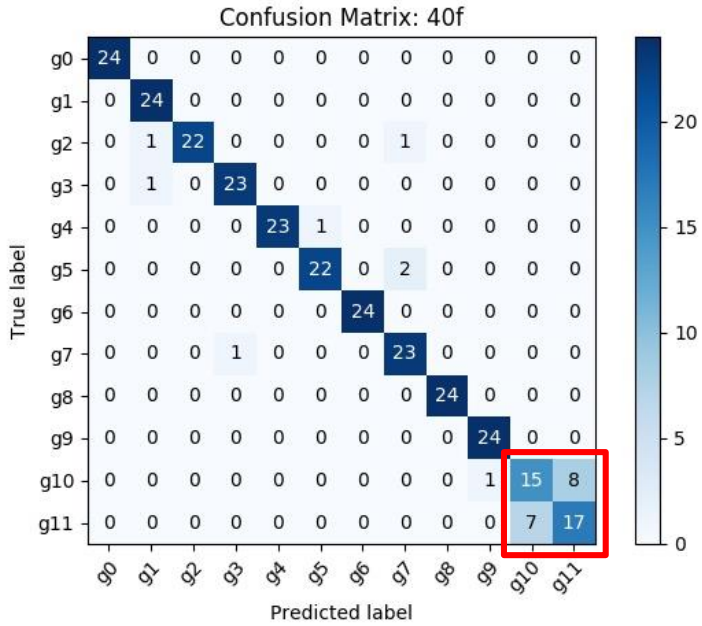
DenseNet

Accuray CNN per tipo di input				
Seq. len.	Depth map	IR	RGB	RGB grayscale
40	**0,903**	0,861	0,837	**0,871**
35	0,903	0,878	**0,851**	0,865
30	0,861	**0,892**	0,837	0,851



Architettura Multimodale

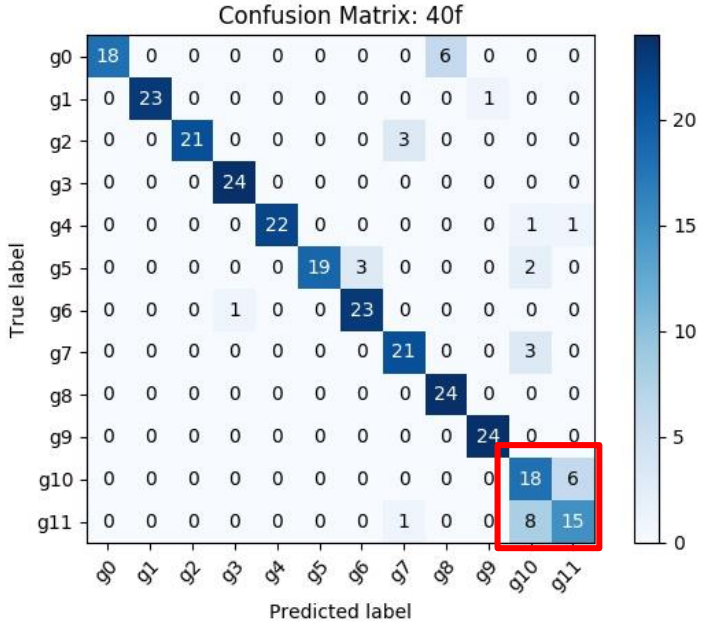
Accuracy per modalità di input	
Input	Accuracy
Depth Map, IR	**0,920**
Depth Map, RGB	0,896
Depth Map, RGB Grayscale	0,910
IR, RGB	0,865
IR, RGB Grayscale	0,865
Depth Map, IR, RGB	0,910
Depth Map, IR, RGB Grayscale	0,899



Risultati Sperimentali - 2

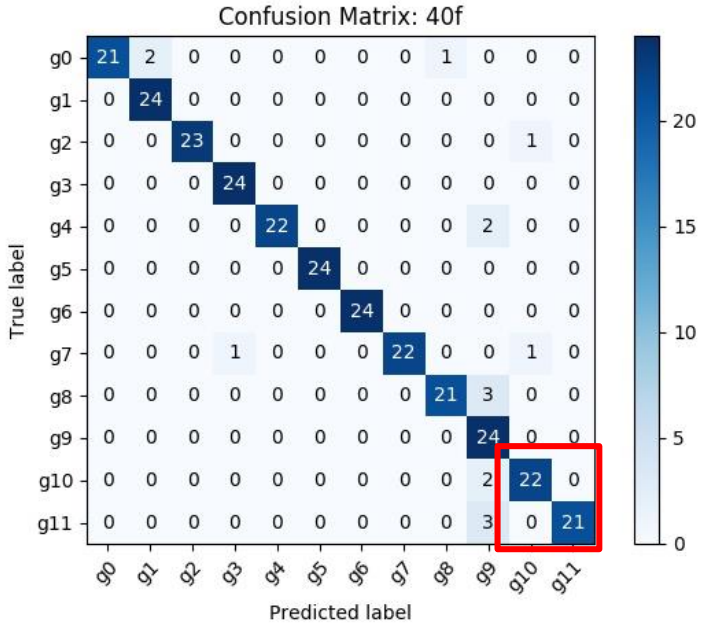
C3D

Accuracy C3D per tipo di input	
Input	Accuracy
Depth Map	0,760
Immagini IR	**0,875**
Immagini RGB	0,722
Immagini RGB Grayscale	0,633



LSTM

LSTM - 3D Features	
Seq. len.	Accuracy
40	**0.944**
35	0.927
30	0.927



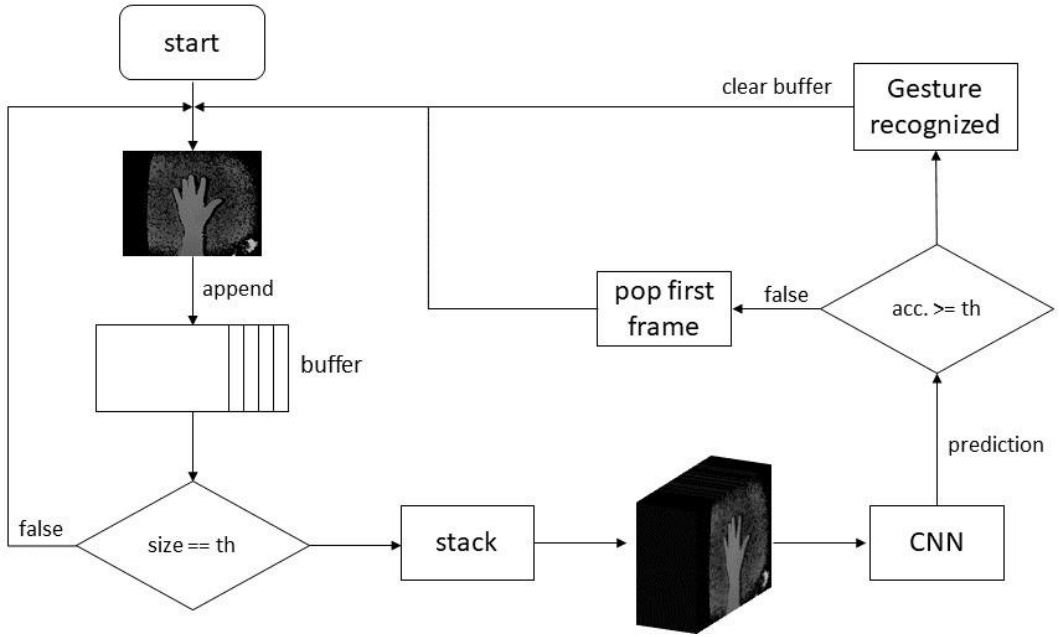
Metodo Implementato

Accuracy			
CNN (Depth Map - 40)	Architettura Multimodale (Depth Map, IR - 40)	C3D (IR - 40)	LSTM (3D features - 40)
0,903	0,920	0,875	**0,944**

CNN – Depth Map

- Affidabilità
- Prestazioni real-time

Prototipo di Sistema



AImage Lab

AImage Lab

Grazie per l'attenzione

Pubblicazioni:

1. 12th Eurographics Workshop on 3D Object Retrieval (3DOR), SHREC 2019 Track: online gesture recognition. F. Manganaro, S. Pini, G. Borghi, R. Vezzani, R. Cucchiara et al.
2. ICIAP 2019, Hand Gestures to Interact with the Car: the xxx dataset. Fabio Manganaro, Stefano Pini, Guido Borghi, Roberto Vezzani, Rita Cucchiara. (SUBMITTED)