



Limiti del riconoscimento dei gesti da singole immagini mediante CNN

Università degli Studi di Salerno
Corso di Laurea Triennale in Informatica



Riconoscimento dei gesti della mano

Nonostante siano uno dei modi più espressivi per comunicare, i gesti della mano sono ancora poco utilizzati nella progettazione di sistemi interattivi a causa delle varie limitazioni tecnologiche.

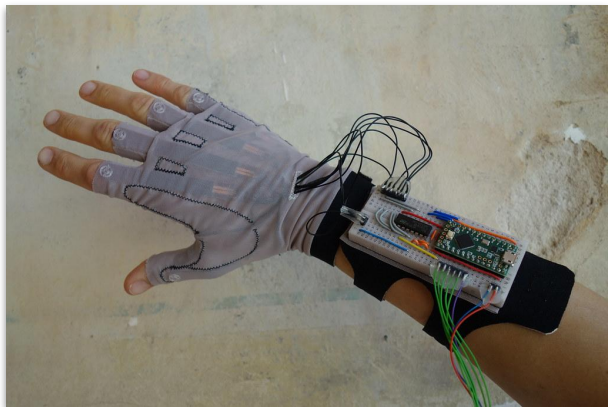
Questi possono essere usati in vari campi, tra cui:

- **sistemi medici e tecnologie per l'assistenza sanitaria;**
- **videogiochi e intrattenimento;**
- **interazione uomo-robot.**

Approcci utilizzati

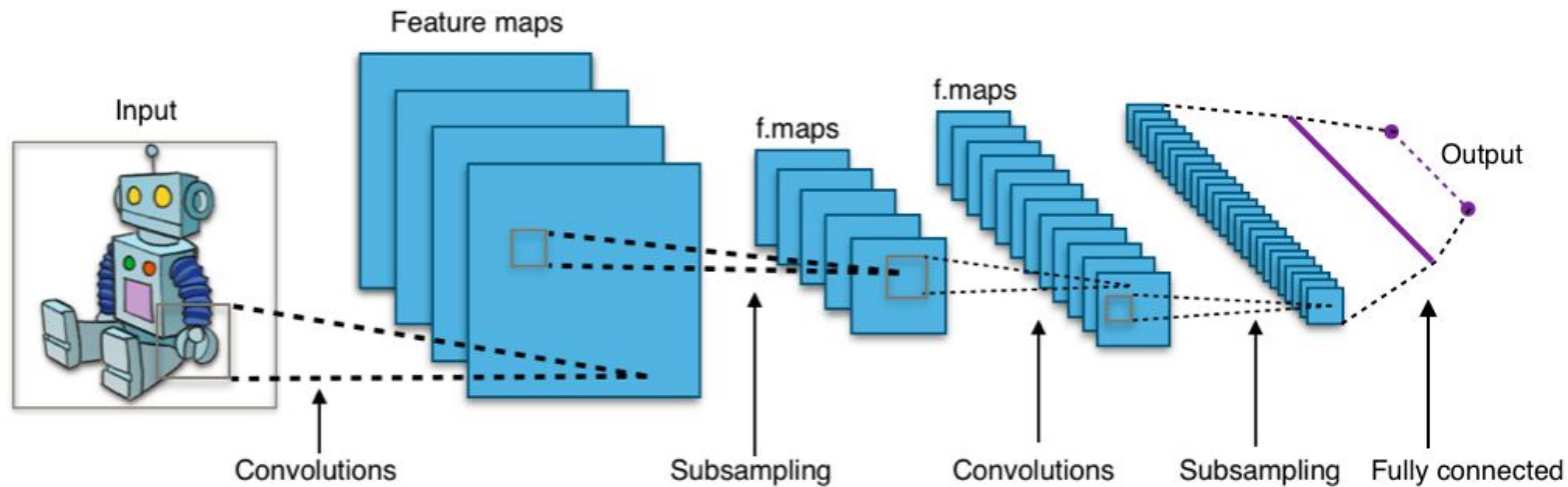
Generalmente sono due gli approcci che vengono impiegati per il riconoscimento dei gesti della mano:

- dataglove;
- tecniche basate sulla visione artificiale.



Reti neurali convoluzionali (CNN)

Le CNN sono uno strumento fondamentale nel campo del deep learning e il loro funzionamento può essere riassunto in due fasi: **estrazione delle caratteristiche** e **classificazione**.



Dataset

Il dataset usato è il Sign Language MNIST:

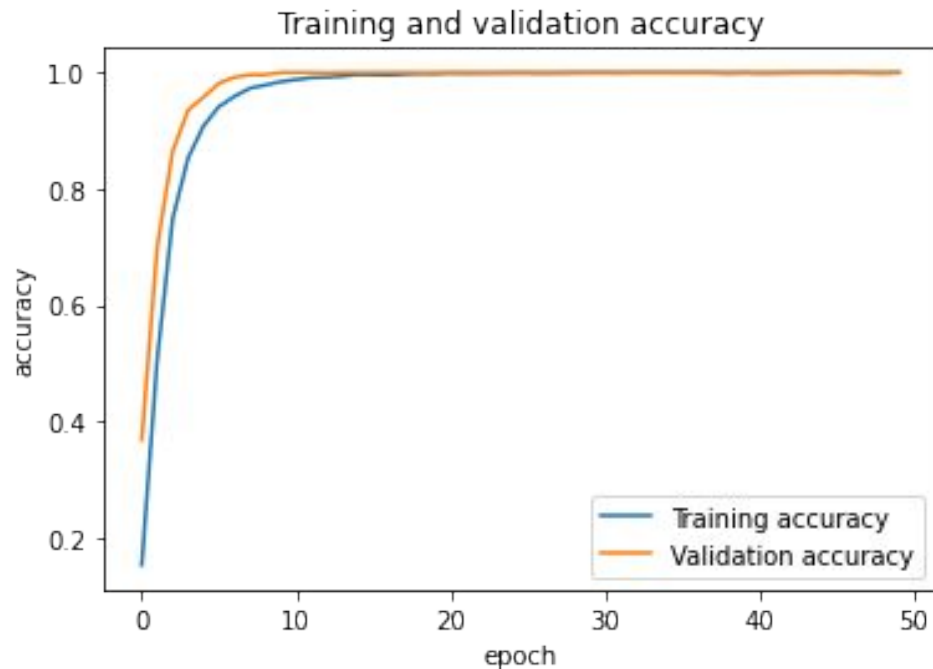
- 34627 immagini di gesti dell'**alfabeto della lingua dei segni americana** (J e Z escluse);
- immagini di dimensioni 28x28x1.



Addestramento

La CNN è stata addestrata utilizzando il linguaggio di programmazione Python insieme alle librerie TensorFlow e Keras.

Questa ha raggiunto un'accuratezza del 100% che però non rispecchia quella reale.





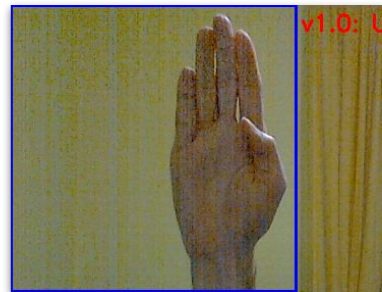
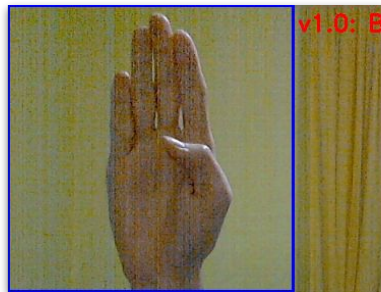
Applicazione desktop

Per testare la CNN è stata realizzata un'applicazione desktop usando il linguaggio di programmazione Python insieme alla libreria OpenCV. Il suo funzionamento è riassumibile nei seguenti step:

1. viene catturato un frame della webcam;
2. viene ritagliata una parte prestabilita del frame;
3. la parte ritagliata viene trasformata in scala di grigi e ridimensionata;
4. la parte ridimensionata viene data in input alla CNN e il risultato viene mostrato all'interno dell'applicazione.

Limiti applicazione desktop

- gesti che non si trova al centro dell'area prestabilita;
- gesti che non occupano l'intera regione.





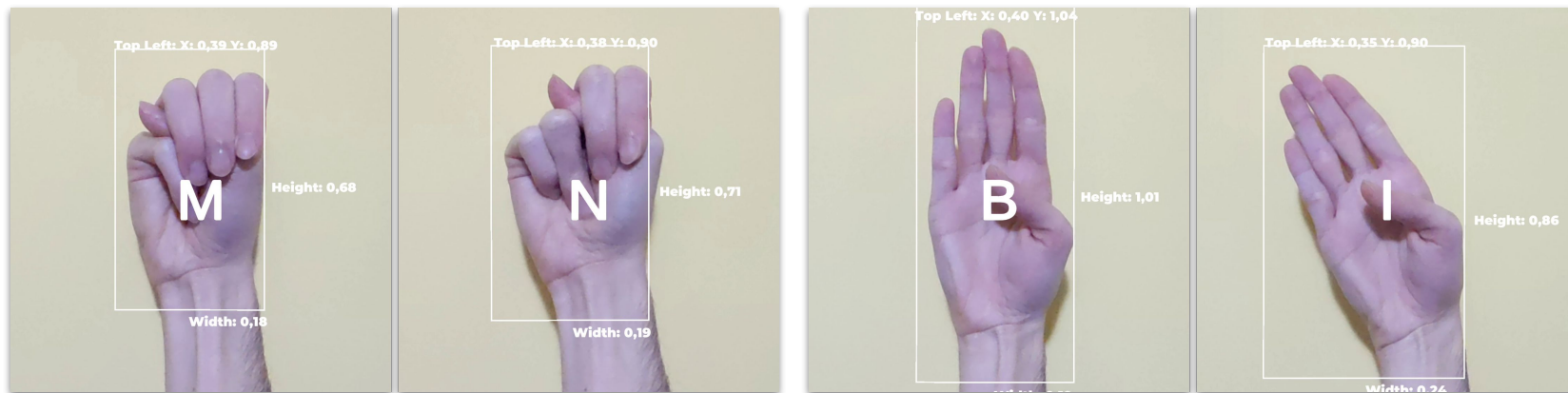
Applicazione Android

Per superare le limitazioni dell'applicazione desktop è stata realizzata un'applicazione Android utilizzando Unity e AR Foundation. Inoltre sono stati implementati il rilevamento e il tracciamento attraverso l'SDK ManoMotion:

- il frame non viene più ritagliato in una regione prestabilita ma in base alla bounding box rilevata dall'SDK.

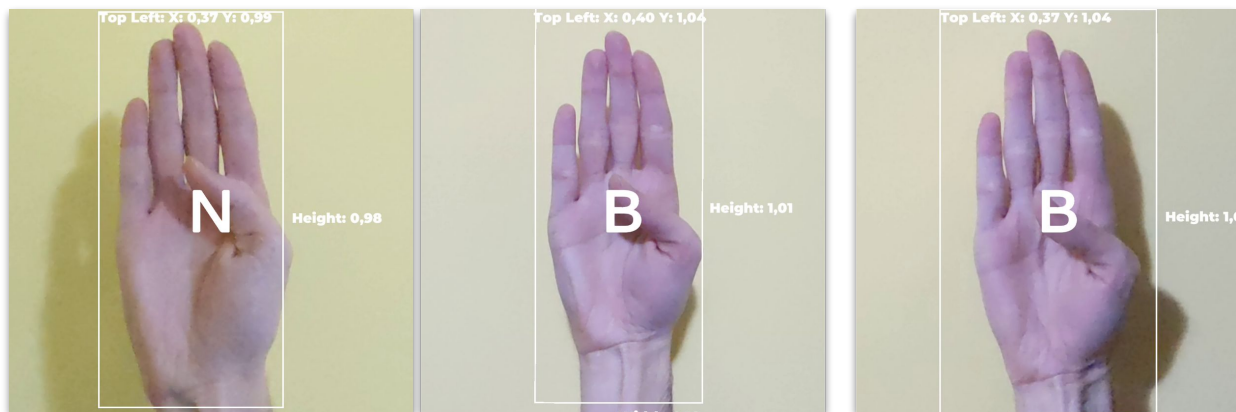
Limitazioni (dataset)

- gesti simili;
- gesti leggermente ruotati.



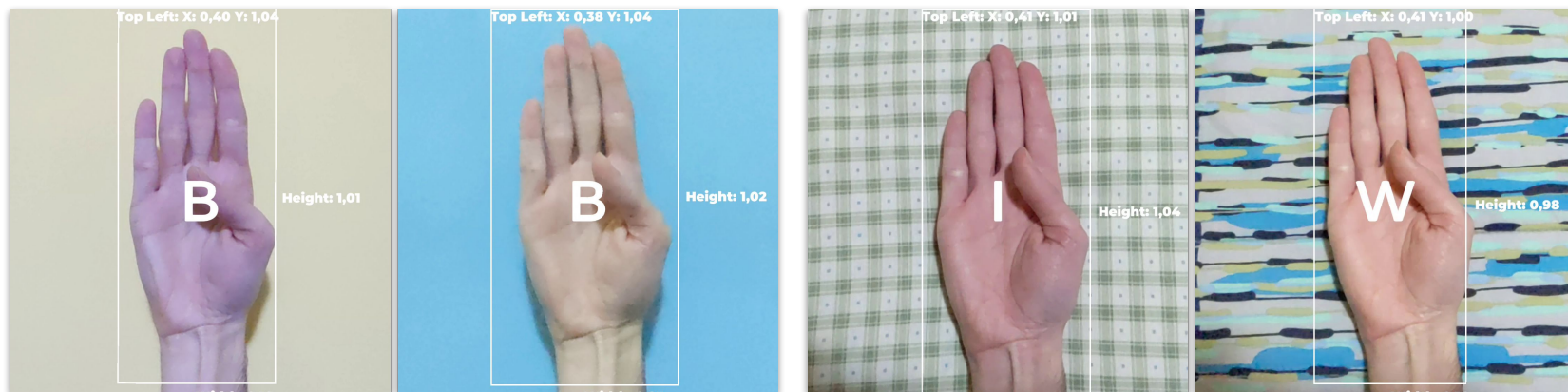
Limitazioni (illuminazione)

- intensità della luce;
- posizione dell'ombra.



Limitazioni (sfondo)

- sfondi complessi.





Conclusioni e sviluppi futuri

- I risultati ottenuti hanno mostrato come l'accuratezza di una CNN non rispecchi sempre quella reale. Da qui la necessità di avere un **benchmark** pubblico (un dataset curato di grandi dimensioni) per confrontare le varie implementazioni.
- Rimane auspicabile approfondire l'efficacia del metodo proposto utilizzando immagini di **dimensioni maggiori** o accompagnate da **informazioni aggiuntive**.