

Università degli Studi di Padova – Dipartimento di Ingegneria dell'Informazione
Corso di Laurea in INGEGNERIA INFORMATICA

Relazione per la prova finale
***«Identificazione e stima del movimento umano
attraverso telecamere ad eventi»***

Relatore: Prof. DAMIANO VARAGNOLO

Laureando: DIEGO MARCHINI
Matricola n.: 2101767

Padova, 21/07/2026

EVENT-BASED CAMERA

Sensore di visione neuromorfica.

Ciascun pixel opera in modo **asincrono** e **indipendente**.

Anziché catturare interi frame a intervalli regolari, ogni pixel reagisce a un cambiamento di intensità luminosa **generando un evento** nel caso tale cambiamento superi una certa soglia.

Ogni **evento** è rappresentato da una tupla

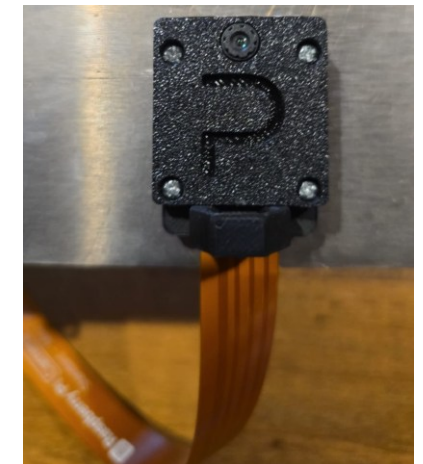
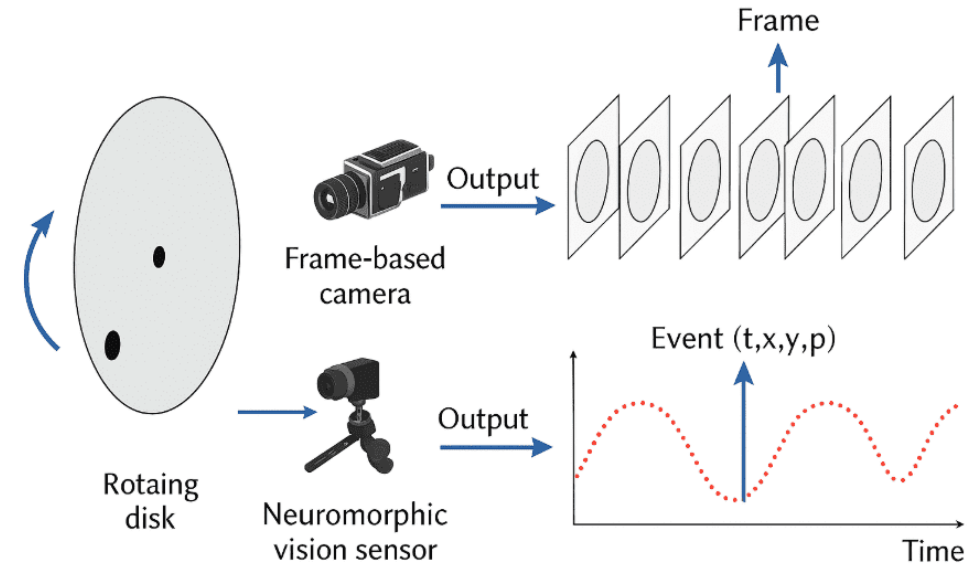
$$e_i = (x_i, y_i, p_i, t_i)$$

dove

x, y : coordinate spaziali del pixel

p : polarità dell'evento (aumento/diminuzione di luminosità)

t : timestamp (quando è avvenuto il cambiamento)



EVENT-BASED CAMERA

VANTAGGI

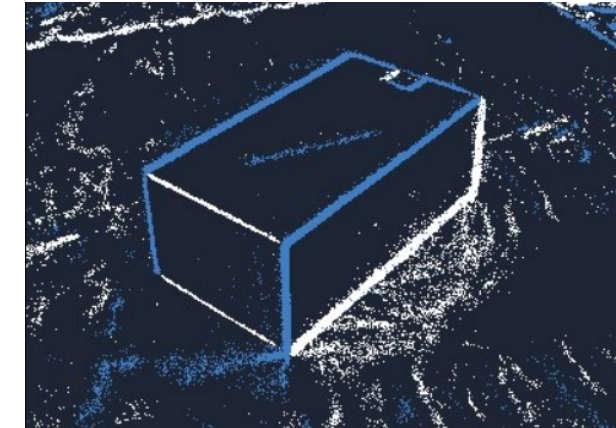
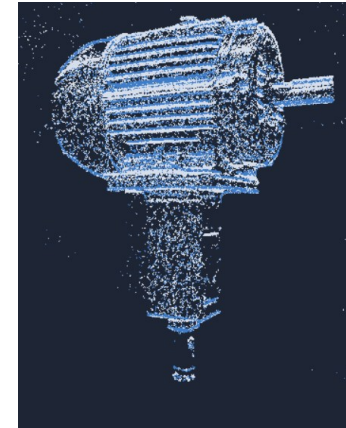
- Elevata risoluzione temporale
- Ampia gamma dinamica (> 120 dB)
- Bassa latenza ($\sim \mu\text{s}$)
- Assenza di motion blur
- Data-volume ridotto
- Basso consumo energetico

LIMITI

- Limitata risoluzione spaziale
- Complessità di elaborazione dati
- Sensibilità al rumore
- Limitazioni lato software
- Costo elevato
- Difficoltà di integrazione con altri sistemi

CASI D'USO

Guida autonoma
Robotica
Automazione industriale
Sistemi di sorveglianza
Realtà virtuale



FRAME VS EVENT

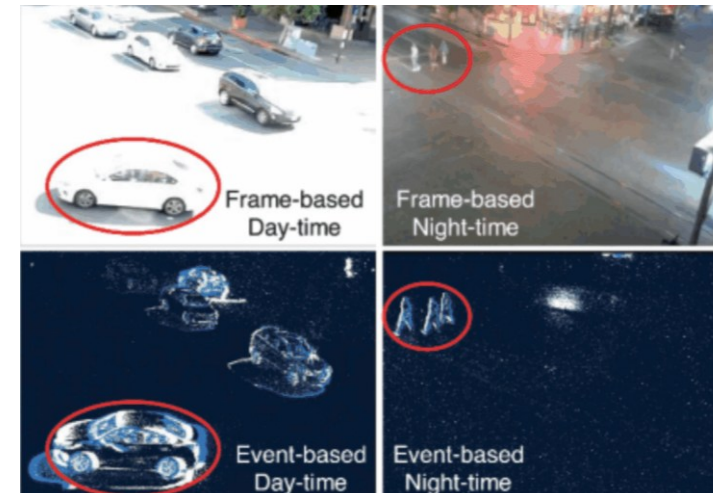
Telecamere tradizionali

- Frame-based
- Funzionamento sincrono
- Frame-rate limitato
- Gamma dinamica: 60-70 dB
- Alta latenza
- Alto consumo
- Data-volume elevato



Telecamere ad eventi

- Event-based
- Funzionamento asincrono
- Alta risoluzione temporale
- Gamma dinamica: >120 dB
- Bassa latenza
- Basso consumo
- Data-volume ridotto



OBIETTIVO DEL LAVORO



Azione: identificare il movimento umano utilizzando sensori di visione neuromorfica, sfruttando i vantaggi di questi ultimi per superare i limiti delle telecamere tradizionali.



Finalità: ottenere dati sulla deambulazione utili per la riabilitazione e la ricerca sul movimento.

Registrazione di una camminata su sfondo statico (minimo disturbo)

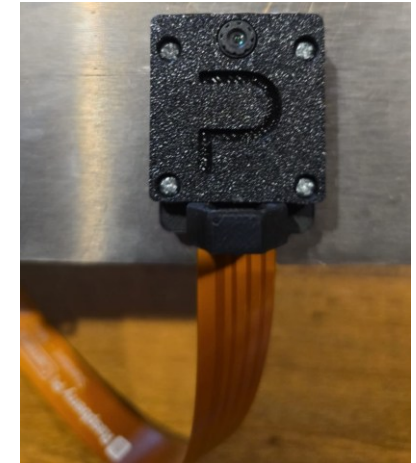
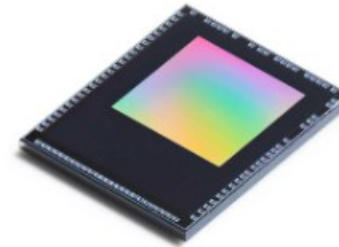
- Utilizzo di un sensore neuromorfico **Metavision GenX320** prodotto da Prophesee:

Risoluzione: 320x320px

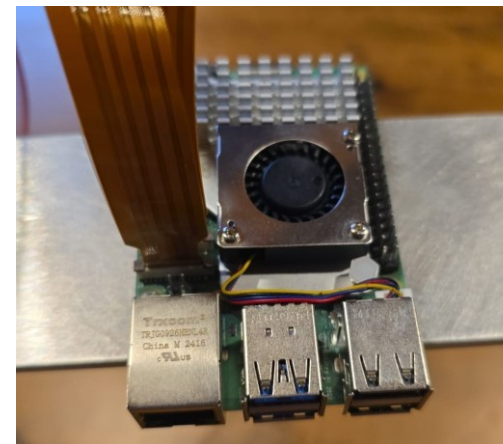
Latenza: $< 150\mu\text{s}$ (1 μs timestamp resolution)

HDR: $> 140\text{dB}$

Low Power (da 3mW fino a 36 μW)



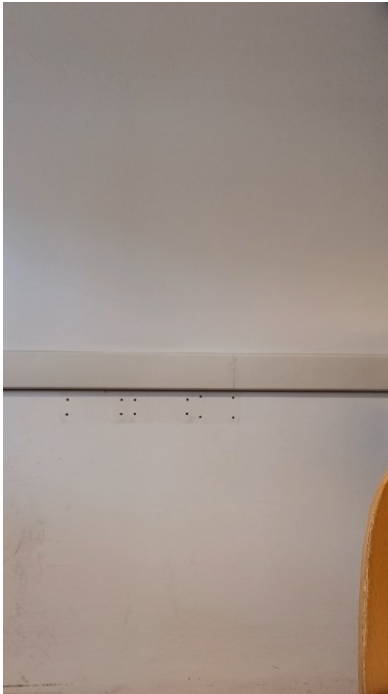
- Utilizzo di un **Raspberry Pi 5** per il backend



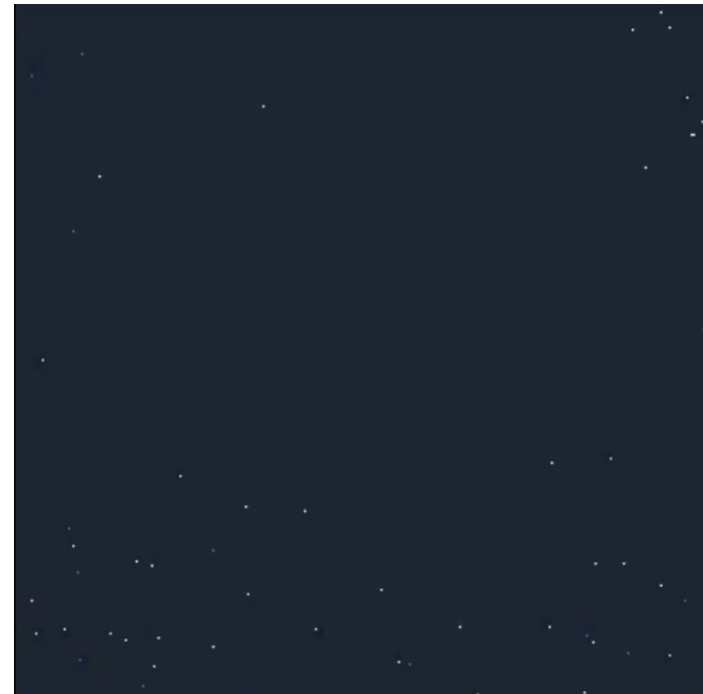
(Ringrazio Danny Tomasi per la collaborazione)

Confronto tra registrazione video e registrazione a eventi convertita in frame (senza elaborazioni)

Video



Event-based frames



Nota: nella registrazione ad eventi (a destra), i colori differenti rappresentano eventi con polarità positiva e negativa

Accumuleremo gli eventi in frame per poi elaborare l'immagine ottenuta, che presenta già una distribuzione di **informazione concentrata** nelle zone di interesse (il soggetto che cammina).



Presenza di rumore (rosso): eventi generati nonostante lo sfondo statico scelto

Mancanza di eventi (verde): alcune regioni «vuote» a causa della ridotta dinamicità della scena

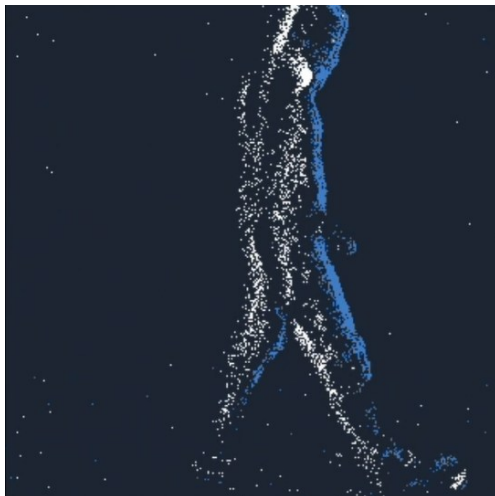
ELABORAZIONE DEI DATI

Idea: ottenere una sagoma ben delimitata per identificare i punti chiave del corpo umano e studiarne il movimento

Per ottenere ciò occorre:

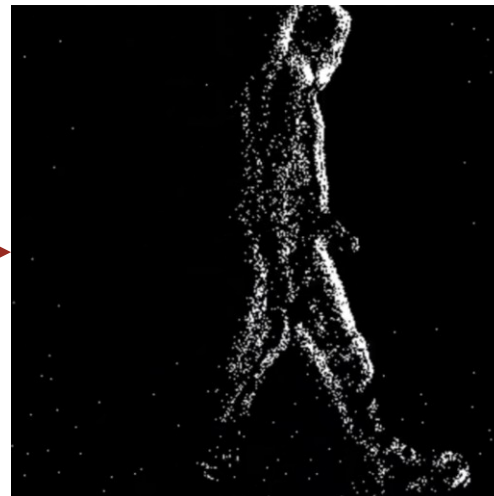
- **binarizzare** il frame
- **filtrare il rumore**
- effettuare operazioni di **dilatazione/riempimento** per coprire la mancanza di eventi

Originale (con polarità)



binarizzazione

Originale (binarizzato)



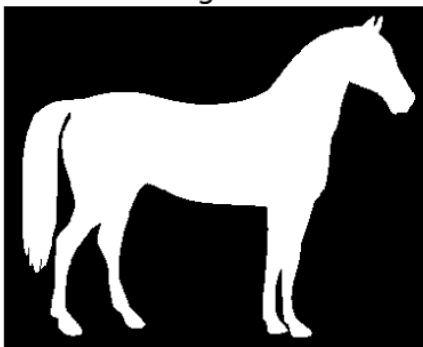
pulizia
+
riempimento

Sagoma

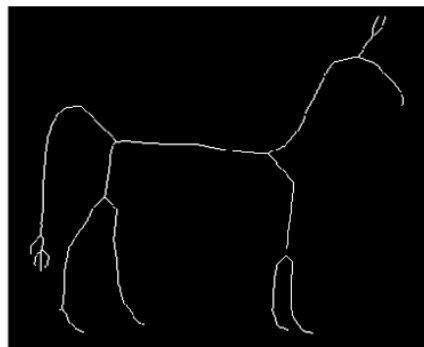


Step aggiuntivo (opzionale): identificazione di una struttura essenziale tramite «scheletrizzazione»

original



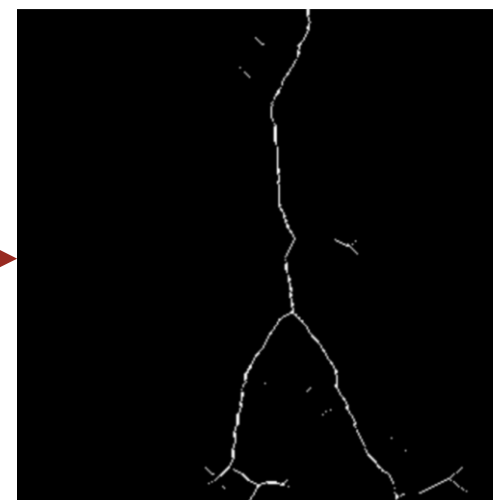
skeleton



Scheletrizzazione: processo con cui si riducono regioni di un'immagine binaria a uno «scheletro» preservando le principali connessioni e scartando la maggioranza dei pixel originali



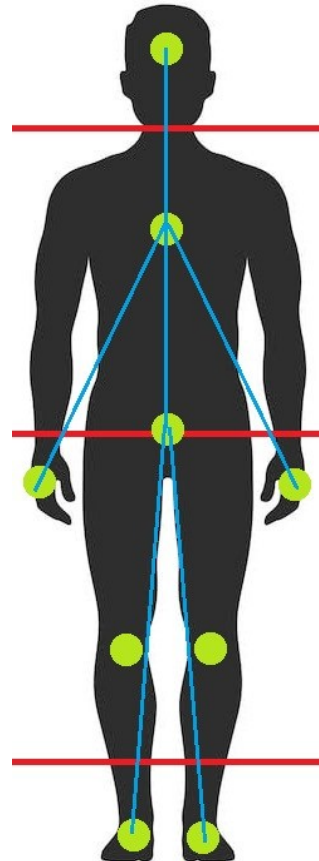
scheletrizzazione



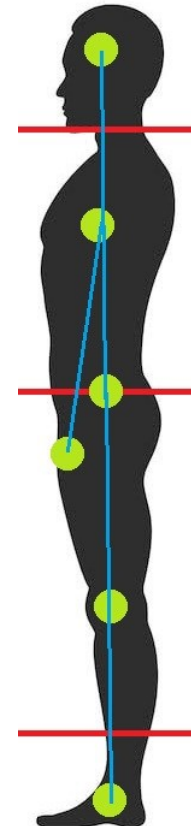
IDENTIFICAZIONE DEL MOVIMENTO

Approccio geometrico: identifichiamo i punti chiave del corpo sfruttando la geometria della figura umana

Vista frontale

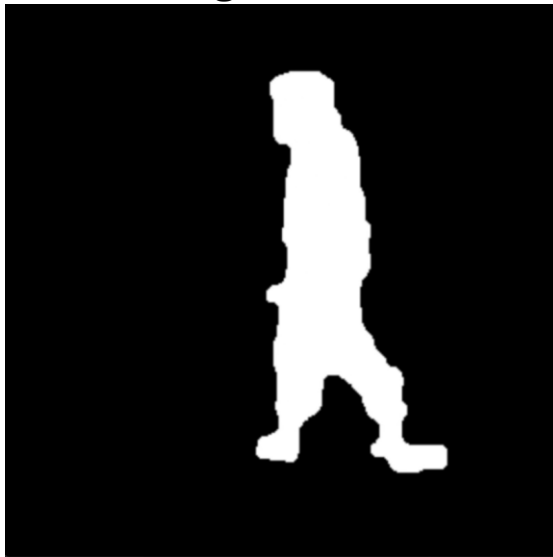


Vista laterale



A partire dalla sagoma «pulita» ricavata dalla registrazione ad eventi, identifichiamo i punti chiave del corpo e li sovrapponiamo alla registrazione originale, ottenendo **dati sul movimento** e mantenendo una visualizzazione con **informazione aggiuntiva minima**.

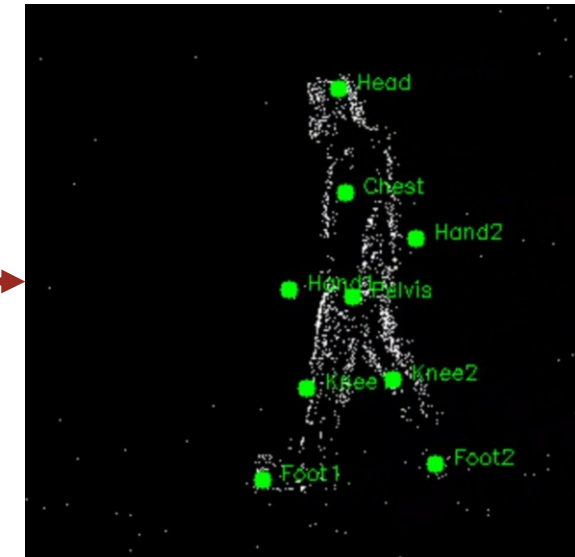
Sagoma



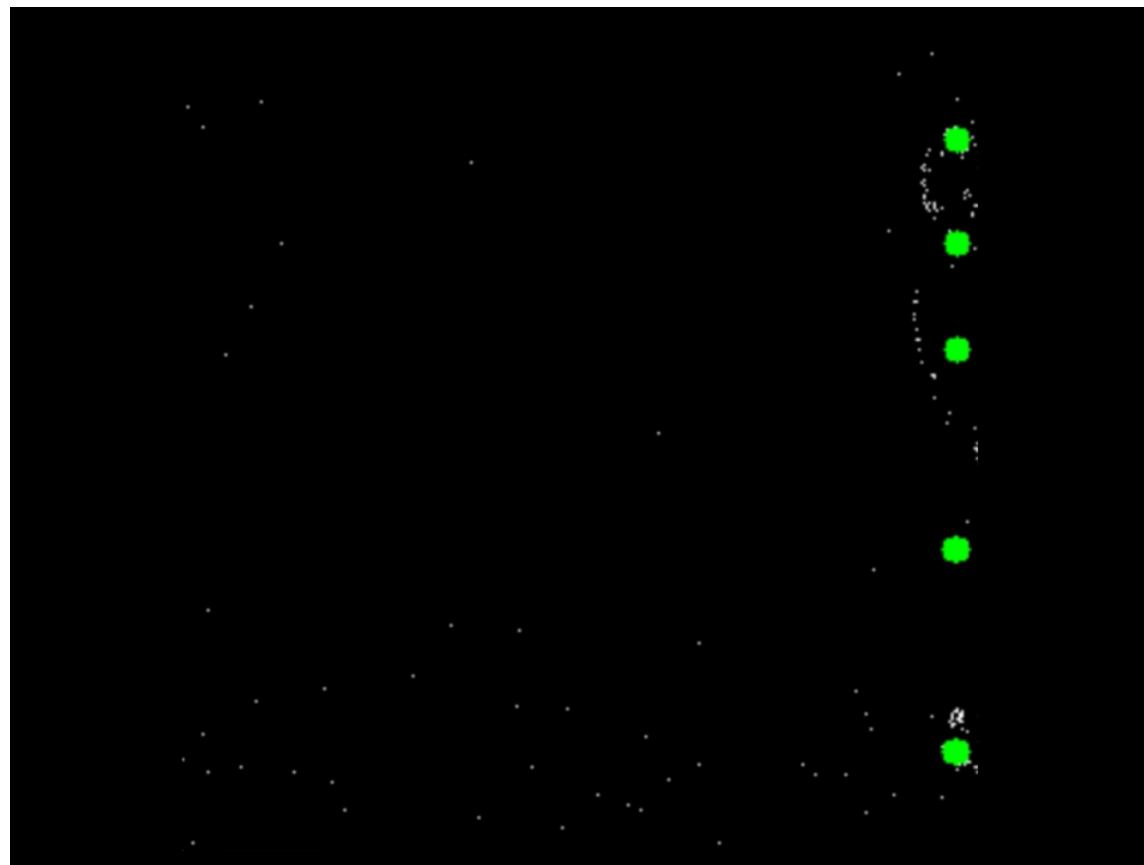
Calcolo dei landmark



Originale + landmark



Originale + landmark



Video con accumulo di eventi con $\text{delta}=2000\mu\text{s}$

SFIDE RISCONTRATE

Curva di apprendimento software: complessità iniziale nell'utilizzo di framework specifici (MetavisionSDK/OpenEB)

Gestione del dato ad eventi: difficoltà nella manipolazione ed elaborazione di un formato di dati non convenzionale

Carenza di librerie di supporto: notevole divario nell'ecosistema software rispetto agli strumenti della computer vision tradizionale

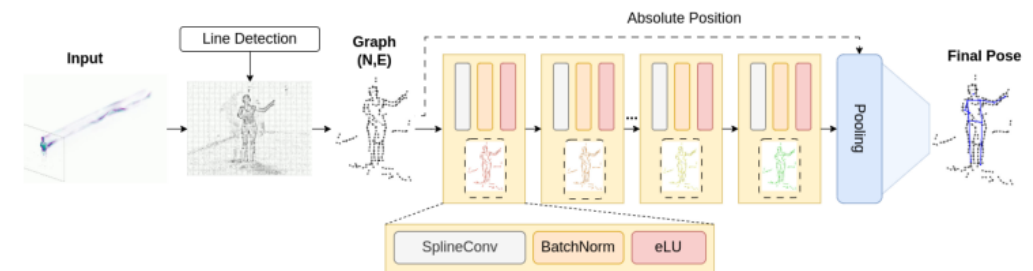
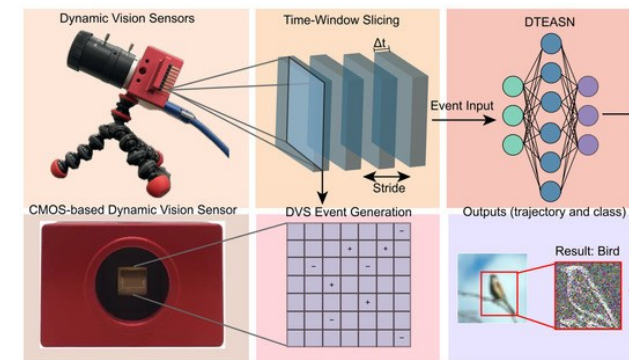
Limitazioni hardware e qualità del dato: il sensore usato (low-power/embedded) ha una risoluzione ridotta (320x320px) rispetto a telecamere tradizionali.

L'approccio usato sfrutta la sparsità degli eventi per ottenere dati «più puliti», ma mantenendo l'elaborazione a livello di frame non riduce appieno il carico computazionale.

Lo sviluppo di algoritmi efficienti per la manipolazione di dati ad eventi è il necessario passo avanti per migliorare le prestazioni di applicazioni real-time.

Proposte per il futuro:

- Impiego di **Spiking Neural Networks (SNN)**
- Impiego di **Graph Neural Networks (GNN)**



GRAZIE PER LA VOSTRA ATTENZIONE