



University of Salerno



Dept. of Information Eng., Electrical Eng. and
Applied Mathematics University of Salerno, Italy

Master's Degree Course in Computer Engineering

Artificial Vision Project Work 2025/2026

“Soccer Net: Player Detection,
Tracking and Behavior Analysis,”

Gruppo n. 16

Antonio Apicella – Antonio Graziosi

Academic year 2025-2026

<u>1. Introduzione</u>
<u>1.1 Contesto e obiettivo del progetto</u>
<u>1.2 Strategie di ottimizzazione e scelte progettuali</u>
<u>2. Dataset</u>
<u>2.1 Descrizione dei dati</u>
<u>2.2 Preprocessing delle Annotazioni</u>
<u>2.2 Analisi delle problematiche</u>
<u>3. Proposed Method</u>
<u>3.1 Network Architecture & Detection Strategy</u>
<u>3.1.1 Modello Adottato: YOLO11x</u>
<u>3.1.2 Calibrazione delle Soglie</u>
<u>3.2 Field Masking</u>
<u>3.3 Tracking Algorithm: BoT-SORT e "Pure Motion"</u>
<u>3.3.1 Configurazione "Pure Motion" vs Re-Identification</u>
<u>3.3.2 Ottimizzazione dei Parametri</u>
<u>3.3.3 Analisi Comparativa e Tuning Sperimentale</u>
<u>3.4 Behavior Analysis Methodology</u>
<u>4. Valutazione e Risultati</u>
<u>4.1 Performance Quantitative</u>
<u>Analisi dei Risultati</u>
<u>4.2 Analisi Qualitativa</u>
<u>5. Conclusioni e Sviluppi Futuri</u>
<u>5.1 Sintesi del Lavoro</u>
<u>5.2 Limiti e Migliorie Future</u>
<u>Riferimenti</u>

1. Introduzione

1.1 Contesto e obiettivo del progetto

Il progetto si colloca all'interno della Soccer Video Understanding Benchmark Suite (SoccerNet), un framework di riferimento per la ricerca nell'analisi video automatica applicata al calcio. Il contesto operativo è caratterizzato da clip video di 30 secondi, acquisite da telecamere in movimento con una risoluzione di 1920×1080 pixel. Tali sequenze presentano sfide significative per la Computer Vision, tra cui movimenti rapidi di camera, variazioni di illuminazione e frequenti occlusioni tra giocatori.

L'obiettivo principale del progetto è lo sviluppo di una pipeline di Computer Vision in grado di risolvere due task distinti ma complementari:

- 1. Detection e Tracking:** Il sistema deve rilevare e identificare univocamente le *"persone di interesse"* in campo (include giocatori e arbitri) in tempo reale, mantenendo la consistenza dell'identità attraverso i frame nonostante i movimenti della telecamera e le interazioni complesse in campo.

2. Behaviour Analysis: Parallelamente al tracciamento, il sistema deve stimare il comportamento collettivo, calcolando il numero medio di persone presenti in specifiche aree del campo definite dinamicamente per ogni sequenza.

La qualità del sistema viene valutata mediante una metrica composita, il PTBS (*Player Tracking and Behavior Score*), definita come la somma di due indicatori:

- HOTA@0.5: metrica principale per il tracking, che bilancia l'accuratezza di rilevamento **DetA** e l'accuratezza di associazione temporale **AssA**.
- nMAE: metrica per l'analisi comportamentale, che quantifica l'errore medio normalizzato nella stima del numero di giocatori nelle ROI, premiando la precisione del conteggio.

1.2 Strategie di ottimizzazione e scelte progettuali

In questo progetto, proponiamo una pipeline completa e ottimizzata per l'analisi video calcistica, introducendo diverse soluzioni tecniche mirate a risolvere le criticità specifiche del dominio SoccerNet. I nostri contributi principali possono essere riassunti come segue:

- **Architettura di Tracking robusta comprensiva di GMC**

È stata implementata una pipeline di Tracking-by-Detection che integra il rilevatore YOLO11 con l'algoritmo di tracciamento BoT-SORT. Per mitigare gli errori causati dai rapidi movimenti di camera, tipici delle riprese televisive, è stato configurato il tracker utilizzando la compensazione del movimento globale (GMC) basata su Sparse Optical Flow. Questa scelta architettonica stabilizza le predizioni del filtro di Kalman, garantendo traiettorie consistenti anche quando i giocatori escono temporaneamente dall'inquadratura.

- **Maschera del campo da gioco**

Per affrontare il problema dei falsi positivi generati da staff a bordo campo, è stato sviluppato un algoritmo di segmentazione adattiva del campo da gioco. Il modulo combina l'analisi dell'istogramma nel canale spazio colore HSV con il filtraggio nel canale spazio colore LAB per identificare dinamicamente la tonalità del manto erboso, adattandosi alle variazioni di illuminazione. Una fase di post-processing morfologico e l'applicazione della Convex Hull assicurano una maschera binaria pulita, utilizzata per scartare automaticamente le detection esterne all'area di gioco.

- **Metodologia di analisi comportamentale e tuning sperimentale**

È stato sviluppato un modulo di analisi comportamentale basato su proiezioni geometriche, che stima la presenza dei giocatori nelle ROI valutando la posizione del punto centrale della base del bounding box, "feet point". Il sistema è stato validato attraverso una campagna di tuning degli iperparametri, in cui abbiamo ottimizzato le soglie di confidenza, i buffer di tracciamento e i parametri di matching per massimizzare la metrica PTBS.

2. Dataset

Il dataset utilizzato per questo progetto è derivato dalla SoccerNet Video Understanding Benchmark Suite. dati consistono in clip video di partite di calcio reali, acquisite da una telecamera principale in movimento che segue l'azione di gioco.

2.1 Descrizione dei dati

Le sequenze video presentano una risoluzione di 1920×1080 pixel. Ogni clip ha una durata di 30 secondi. Il dataset è organizzato seguendo lo standard *MOT Challenge*: ogni sequenza è contenuta in una cartella specifica che ospita le immagini dei singoli frame (sottocartella `\img1`) e i dati di annotazione.

Le annotazioni di Ground Truth (GT) sono fornite in file di testo CSV (`gt.txt`) che seguono rigorosamente il formato a 10 colonne standardizzato per il tracking multi-oggetto. Ogni riga rappresenta un'istanza di un oggetto in un frame specifico e i valori corrispondono ordinatamente a:

```
Frame ID, Track ID, Bounding Box, Confidence Score, Class, Visibility, 3  
D info
```

Con bounding box indicato con il formato `xywh`, Confidence Score impostato a 1 per il ground truth, e le ultime tre colonne (Class, Visibility, 3D info) sono impostate a -1 in quanto non utilizzate per questo specifico benchmark, ma necessarie per compatibilità con il formato MOT20.

2.2 Preprocessing delle Annotazioni

È importante notare che il dataset originale include annotazioni per tutte le entità in movimento, inclusa la palla. Poiché l'obiettivo del progetto è limitato al tracking delle persone (giocatori e arbitri), è stata applicata una fase di preprocessing automatizzata per rimuovere la classe "ball" dalle Ground Truth, evitando che venisse conteggiata erroneamente come un Falso Negativo (FN) o un Falso Positivo (FP) durante la valutazione.

La procedura di pulizia è stata implementata tramite uno script Python dedicato (`src/utils/remove_balls_from_gt.py`) che opera in tre fasi:

1. Per ogni sequenza, lo script analizza il file di configurazione `gameinfo.ini`, che contiene la mappatura semantica delle tracce. In particolare, il file associa ogni `trackletid_X` alla classe dell'oggetto corrispondente (es. `ball;1`).
2. Vengono estratti tutti i Track ID numerici associati alla stringa "ball" all'interno del file di configurazione.
3. Il file `gt.txt` viene riletto riga per riga; tutte le entry il cui secondo campo (Track ID) corrisponde a uno degli ID identificati come "palla" vengono scartate. Il risultato è un file di annotazione pulito contenente esclusivamente le tracce relative alle persone.

2.2 Analisi delle problematiche

Dall'analisi preliminare del dataset sono emerse delle criticità principali che hanno guidato le nostre scelte architettoniche:

- **Small object detection:** i giocatori spesso occupano un'area estremamente ridotta dell'immagine. Questa caratteristica rende inefficaci i detector standard operanti a basse risoluzioni, imponendo l'utilizzo di una risoluzione di input elevata (1088) per preservare le feature spaziali necessarie al rilevamento.
- **Occlusioni e affollamento:** il calcio presenta dinamiche complesse come mischie, esultanze o calci d'angolo, dove i giocatori si sovrappongono parzialmente o totalmente per diversi secondi. Queste occlusioni frequenti causano la perdita temporanea del target visivo, rendendo critico il mantenimento dell'identità aumentando il rischio di ID Switching o frammentazione delle tracce.
- **Alta similarità intra-classe:** in una partita di calcio i compagni di squadra condividono la stessa divisa identica. I modelli di Re-Identification basati su caratteristiche cromatiche globali falliscono sistematicamente in questo contesto.

3. Proposed Method

In questo capitolo descriviamo la pipeline sviluppata per affrontare le sfide della *SoccerNet Benchmark Suite*. L'architettura proposta segue un approccio **Tracking-by-Detection** a stadi, progettato per massimizzare il rilevamento dei giocatori (Recall) nelle prime fasi e delegare la pulizia dei falsi positivi a meccanismi di filtraggio successivi.

La pipeline si articola in tre moduli sequenziali:

1. **Player Detection:** Rilevamento dei giocatori frame-per-frame.
2. **Field Masking:** Filtraggio spaziale per rimuovere staff e pubblico.
3. **Multi-Object Tracking:** Associazione temporale delle identità.
4. **Behavior Analysis:** Stima della presenza dei giocatori nelle ROI.

3.1 Network Architecture & Detection Strategy

Il primo stadio si basa sul rilevamento robusto dei giocatori. Data la natura varia delle (zoom, grandangolo, movimenti rapidi), i target appaiono spesso piccoli, sfocati ("motion blur") o parzialmente occlusi.

3.1.1 Modello Adottato: YOLO11x

È stato selezionato **YOLO11x**¹ (versione extra-large) come detector principale, optando per l'utilizzo dei pesi pre-addestrati sul dataset COCO senza effettuare un ulteriore fine-tuning sul dataset SoccerNet. Questa decisione è stata presa dopo aver osservato che il modello base possiede già una capacità di generalizzazione superiore rispetto a modelli specifici ma più piccoli. In particolare, YOLO11x si è dimostrato eccezionalmente robusto nell'individuare persone in pose non convenzionali o articolate (es. scivolate, salti, cadute), situazioni estremamente frequenti nel gioco del calcio ma spesso critiche per detector meno potenti.

La scelta della variante Extra-Large è motivata dalla necessità di massimizzare la capacità di estrazione delle feature su oggetti di piccole dimensioni. Nelle inquadrature in cui viene ripreso l'intero campo di gioco, i giocatori occupano pochi pixel e modelli più leggeri hanno mostrato limiti evidenti.

In fase preliminare, infatti, è stato testato un modello **YOLOv8n fine-tuned**² specificamente sul dataset SoccerNet. Nonostante l'addestramento specifico, questo modello faticava a mantenere stabili le detection dei calciatori più lontani o parzialmente occlusi, frammentando le tracce e abbassando la metrica DetA. Al contrario, YOLO11x ha garantito un Recall nettamente superiore sui piccoli oggetti, rendendo superfluo, se non controproducente in termini di tempi/benefici, un ulteriore addestramento specifico.

Il modello è stato configurato con una risoluzione di input di **1088 pixel**. Questa dimensione, multiplo di 32, rappresenta un buon trade-off tra la preservazione dei dettagli spaziali e il costo computazionale, evitando l'upsampling eccessivo che avrebbe introdotto artefatti senza reali benefici informativi.

3.1.2 Calibrazione delle Soglie

Contrariamente alle configurazioni standard che privilegiano la precisione, è stata adottata una strategia aggressiva orientata al *Recall*. L'assunto di base è che il tracker può filtrare il rumore, ma non può recuperare un giocatore mai rilevato.

- **Confidence Threshold (`conf: 0.2`):** La soglia di confidenza è stata impostata a 0.2, un valore molto basso scelto per minimizzare il rischio di perdere detection valide (Falsi Negativi). Dato che il detector ha un'alta confidenza sulla classe "persona" (*pedestrian*) e il contesto è limitato al campo di gioco, il rischio di confondere i giocatori con altri oggetti è minimo.
Questa configurazione ci permette di rilevare il maggior numero possibile di soggetti, inclusi quelli sfocati dal movimento ("motion blur"), occlusi o lontani dalla telecamera. Nonostante l'aumento dei Falsi Positivi iniziali, questa strategia è essenziale per mantenere le tracce stabili nel tempo ed evitare che le traiettorie si spezzino durante le azioni veloci.
- **IoU Threshold (`iou: 0.7`):** La soglia per la *Non-Maximum Suppression* (NMS) è stata fissata a 0.7. Un valore alto è necessario per gestire le situazioni di affollamento (es. calci d'angolo, mischie), dove i bounding box dei giocatori si sovrappongono significativamente. Una NMS più bassa avrebbe soppresso erroneamente giocatori vicini.

3.2 Field Masking

Un problema critico riscontrato nel dataset SoccerNet è la presenza di persone "non giocanti" (fotografi, steward, pubblico nelle prime file) che vengono correttamente rilevate dal detector come classe "persona", ma che costituiscono Falsi Positivi (FP) per il task di tracking dei giocatori.

Per mitigare questo problema, abbiamo implementato un modulo di Field Masking (`src/utils/field_masking.py`) che agisce come filtro spaziale post-detection: un rilevamento viene accettato solo se il punto centrale della base del suo bounding box ("feet point") cade all'interno dell'area di gioco segmentata.

L'algoritmo di segmentazione sviluppato non si basa su ROI statiche, ma è progettato per essere adattivo e robusto a variazioni di illuminazione. La pipeline di elaborazione si articola in cinque fasi sequenziali:

1. Preprocessing e Downscaling:

L'immagine viene inizialmente ridimensionata con un fattore di scala 0.5 e sottoposta a una sfocatura Gaussiana (5×5). Questo passaggio riduce il rumore ad alta frequenza tipico della texture dell'erba.

2. Analisi Colore Adattiva (Spazio HSV):

Invece di utilizzare soglie fisse ("hard-coded"), l'algoritmo calcola l'istogramma del canale Hue (Tonalità) concentrando esclusivamente sul 60% inferiore dell'immagine, dove è statisticamente probabile trovare il terreno di gioco. Viene identificato il picco modale dell'istogramma (il colore dominante) e definito un range dinamico di tolleranza (± 18 valori

di Hue, con saturazione > 35 e valore > 30). Questo permette alla maschera di adattarsi automaticamente alla tonalità specifica del prato di ogni stadio.

3. Raffinamento Semantico (Spazio LAB):

Parallelamente, l'immagine viene convertita nello spazio colore CIELAB. Utilizziamo specificamente il canale 'a' (asse verde-rosso), imponendo una soglia empirica ($a < 118$) per isolare le componenti cromatiche verdi. La maschera preliminare è ottenuta dall'intersezione logica (`bitwise_and`) tra la maschera HSV (adattiva) e la maschera LAB (robusta alla luminosità), eliminando efficacemente falsi positivi cromatici.

4. Pulizia Morfologica:

Alla maschera binaria vengono applicate operazioni morfologiche con kernel ellittici:

- **Opening (15 × 15):** Rimuove piccoli blob isolati di rumore (es. coriandoli, scarpini colorati).
- **Closing (25 × 25):** Riempie i "buchi" all'interno della maschera causati dalle linee bianche del campo o dalle figure dei giocatori stessi.

I kernel ellittici approssimano una forma circolare, questa evita l'introduzione di artefatti geometrici squadrati ("aliasing" sugli angoli) che un kernel rettangolare creerebbe. Inoltre, poiché la figura umana proiettata sul piano immagine tende ad avere contorni curvi e organici, l'elemento strutturante ellittico preserva meglio la naturalezza dei bordi rispetto a forme rigide.

5. Regolarizzazione Geometrica (Convex Hull):

L'ultimo step sfrutta la proprietà geometrica dei campi da calcio di essere aree convesse. Viene estratto il contorno con l'area maggiore e su di esso viene calcolata la Convex Hull. Questo passaggio è fondamentale per includere nella maschera eventuali zone interne non rilevate e per regolarizzare i bordi frastagliati, ottenendo un poligono solido e coerente. La Convex Hull è il più piccolo poligono convesso che racchiude completamente un insieme di punti, poiché la porzione visibile di un campo da calcio è approssimabile a un'area convessa, questa operazione è utile per "sanare" le irregolarità sui bordi. Se, ad esempio, la presenza di giocatori o cartelloni a bordo campo interrompe la continuità del verde creando delle insenature nella maschera, la Convex Hull riempie automaticamente questi vuoti.

Infine, la logica di validazione (`is_point_on_field()`) include una tolleranza di **40 pixel** sul bordo inferiore, per gestire i casi in cui i piedi del giocatore, a causa della prospettiva della telecamera, appaiono leggermente al di sotto del limite visivo dell'erba.

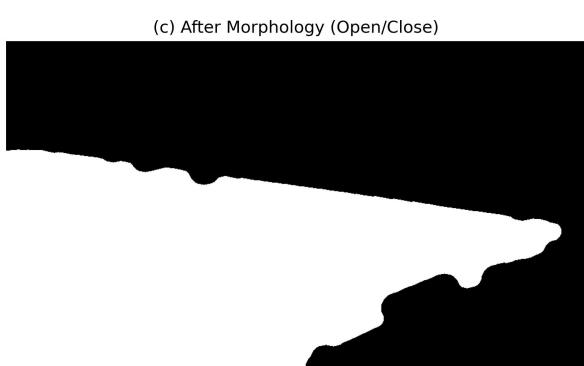
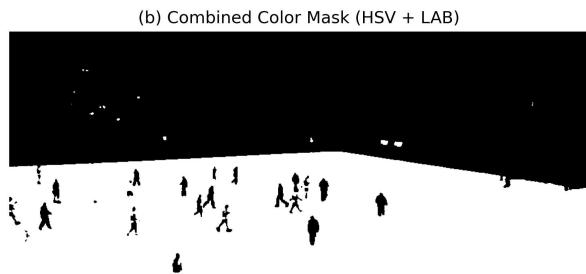
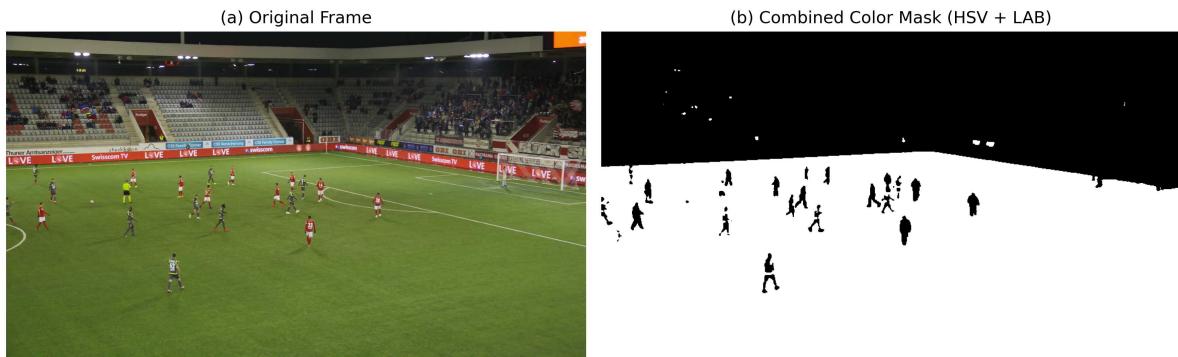


Figura 1: Visualizzazione degli stadi del Field Masking in condizioni differenti. (a) Frame Originale. (b) Maschera Colore Combinata (HSV + LAB). (c) Maschera dopo Operazioni Morfologiche. (d) Maschera Finale con Convex Hull.

3.3 Tracking Algorithm: BoT-SORT e "Pure Motion"

Per l'associazione temporale abbiamo utilizzato **BoT-SORT**, preferendolo a DeepSORT o ByteTrack standard per la sua capacità di gestire i movimenti di camera tramite **GMC (Global Motion Compensation)**.

3.3.1 Configurazione "Pure Motion" vs Re-Identification

Uno dei risultati sperimentali più rilevanti è stata la scelta di disabilitare il modulo di Re-Identification (ReID).

I tentativi iniziali con ReID (anche utilizzando modelli progettati per il ReID come Omni-Scale Network finetuned³ sul dataset della competizione) hanno mostrato un degrado della metrica di associazione (AssA). In un contesto calcistico:

1. I giocatori della stessa squadra indossano divise identiche, rendendo le feature visive ambigue.
2. La bassa risoluzione dei crop dei giocatori (dovuta alla distanza della camera) rende gli embedding poco discriminativi.

Abbiamo quindi optato per una configurazione **Pure Motion**, affidando l'associazione esclusivamente alla sovrapposizione spaziale (IoU) e alla compensazione del movimento della camera (Sparse Optical Flow).

3.3.2 Ottimizzazione dei Parametri

La configurazione finale è il risultato di un tuning volto a bilanciare la stabilità delle tracce (AssA) con la capacità di rilevamento (DetA).

- **Matching Threshold** (`match_thresh: 0.9`): Questa soglia indica la massima distanza IoU (1 – IoU) accettabile per considerare valido un match. Un valore così elevato significa che è sufficiente una sovrapposizione minima del 10% tra la predizione del *Kalman Filter* (corretta via GMC) e la nuova detection per confermare l'associazione. Questa scelta si è rivelata cruciale per il calcio: i movimenti rapidi e i cambi di direzione improvvisi spesso riducono drasticamente l'IoU tra frame consecutivi. Una soglia più severa (es. 0.6) avrebbe spezzato le tracce di giocatori in corsa, mentre questa "tolleranza" permette al tracker di mantenere l'identità affidandosi alla bontà della stima di movimento del Kalman Filter.
- **Track High Threshold** (`track_high_thresh: 0.35`): Questo è il parametro chiave per la DetA. Permettiamo al tracker di aggiornare le traiettorie esistenti anche con detection a bassa confidenza (35%). Se avessimo usato valori standard (0.5 o 0.6), avremmo perso il tracciamento di giocatori in corsa veloce o sfocati, frammentando le tracce.
- **New Track Threshold** (`new_track_thresh: 0.4`): Per bilanciare l'apertura delle soglie di detection, si è stati leggermente più selettivi sulla creazione di *nuove* tracce. Unito al Field Masking, impedisce al "rumore" di generare tracce fantasma.
- **Track Low Threshold** (`track_low_thresh: 0.2`): Le detection sotto questa soglia vengono scartate definitivamente. Questo pulisce le "code" delle tracce, evitando che un ID rimanga

"aggrappato" a detection spazzatura prima di scomparire, migliorando la precisione dei Bounding Box.

- **Track Buffer** (`track_buffer: 150`): Buffer esteso a circa 7 secondi (a 25fps) per gestire occlusioni prolungate o uscite temporanee dal campo.

3.3.3 Analisi Comparativa e Tuning Sperimentale

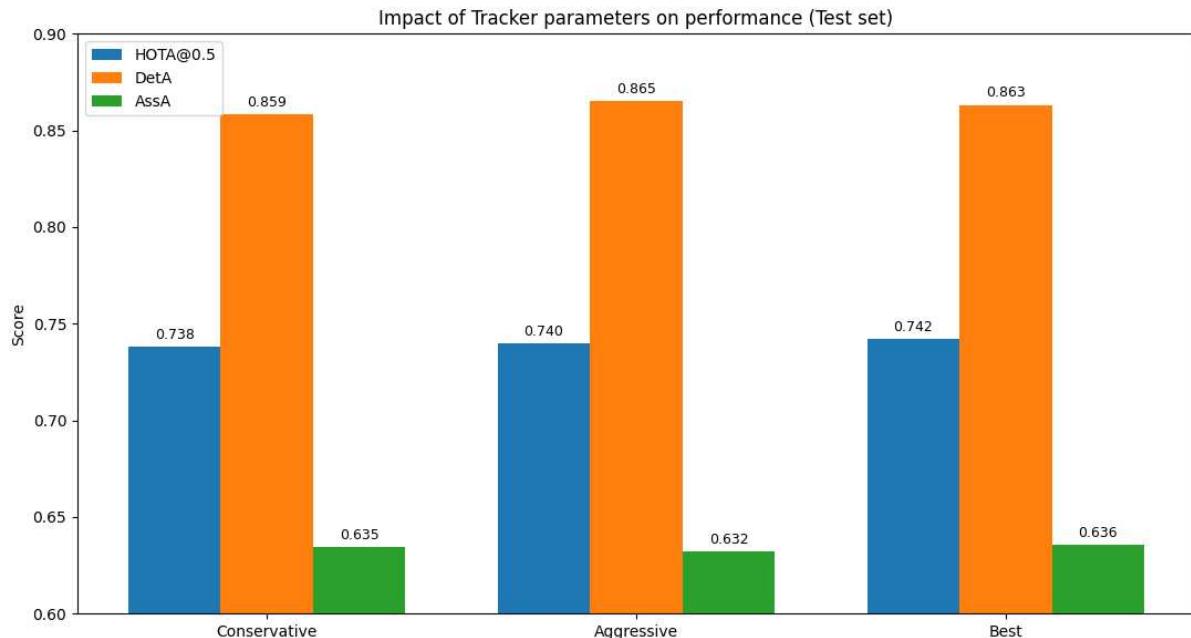


Figura 2: Confronto dell'impatto dei parametri del tracker sulle metriche HOTA, DetA e AssA sul Test Set. Si nota come la configurazione Best massimizzi sia la detection che l'associazione rispetto agli estremi Conservativo e Aggressivo.

Per validare la robustezza della configurazione scelta, è state condotta un'analisi comparativa con diverse varianti dei parametri di tracciamento. In questo paragrafo vengono presentati i risultati ottenuti sul test set su tre esperimenti significativi, riassunti nella [Figura 2](#), evidenziano le sfide affrontate nel contesto calcistico del contest.

1. Configurazione "Conservativa":

- *Parametri:* `new_track_thresh: 0.65`, `track_high_thresh: 0.5`.

Questa configurazione imponeva standard di ingresso molto severi. Sebbene abbia prodotto un'Associazione pulita (**AssA 0.635**) e pochi Falsi Positivi, ha penalizzato pesantemente la Detection Accuracy (**DetA 0.858**). L'analisi dei log ha rivelato un numero eccessivo di **Falsi Negativi (> 50.000)**: molti giocatori, pur essendo rilevati dal detector con confidenza media (es. 0.5-0.6), venivano scartati dal tracker, frammentando le traiettorie e abbassando l'HOTA complessivo a **0.738**.

2. Configurazione "Aggressiva":

- *Parametri:* `new_track_thresh: 0.40`.

Al contrario, abbassando la soglia di ingresso a 0.40 per massimizzare il recupero dei giocatori, è stata osservata un'esplosione dei Falsi Positivi (**circa 29.000**). Nonostante un numero di True Positives superiore persino al miglior risultato finale, l'inclusione di box "spazzatura" ha degradato la precisione spaziale e l'Associazione (**AssA 0.632**), portando l'HOTA a **0.739**.

3. Configurazione “Best”:

- *Parametri:* `new_track_thresh: 0.50`, `track_low_thresh: 0.2`.

La configurazione finale si posiziona esattamente nel punto di equilibrio. Con una soglia di 0.5, recuperiamo i giocatori persi dalla configurazione conservativa senza imbarcare il rumore di quella aggressiva. Inoltre, il mantenimento di un `track_low_thresh` basso (0.2) permette di non troncare prematuramente le tracce dei giocatori che subiscono occlusioni momentanee, garantendo la **DetA** (0.863) e l'**AssA** (0.636) più alte simultaneamente, massimizzando l'HOTA a **0.742**.

3.4 Behavior Analysis Methodology

Il modulo di analisi comportamentale stima il numero di giocatori in due Region of Interest (ROI) definite per ogni video.

L'approccio implementato in `behaviour_analyzer.py` si basa su una logica geometrica:

1. Vengono caricate le coordinate relative delle ROI dal file `roi_config.json`.
2. Per ogni frame, si calcola il "**feet point**" di ogni giocatore tracciato (punto medio del lato inferiore del bounding box).
3. Un giocatore viene contato nella ROI se e solo se il suo feet point è contenuto nel rettangolo della ROI.

Il risultato viene valutato tramite la metrica nMAE (Normalized Mean Absolute Error).

4. Valutazione e Risultati

In questa sezione presentiamo i risultati quantitativi ottenuti sul **Test Set** della SoccerNet Challenge, analizzando le metriche di performance e discutendo i punti di forza emersi dalla validazione sperimentale.

4.1 Performance Quantitative

La tabella seguente riassume le metriche finali ottenute con la configurazione "Best".

Metrica	Valore	Analisi
HOTA	0.742	<i>Higher Order Tracking Accuracy.</i> Il risultato riflette l'ottimo bilanciamento raggiunto tra precisione di rilevamento e consistenza delle traiettorie.
DetA	0.863	<i>Detection Accuracy.</i> Valore particolarmente elevato, testimonianza dell'efficacia della strategia "High Recall" con YOLO11x e soglie basse.
AssA	0.636	<i>Association Accuracy.</i> Un buon risultato considerando l'assenza di un modulo di Re-Identification, ottenuto grazie al tuning fine del <i>Kalman Filter</i> e del GMC.

Analisi dei Risultati

L'analisi dei dati evidenzia come il sistema eccella particolarmente nella **Detection (DetA 0.863)**. La scelta di utilizzare un modello *Extra-Large* con input ad alta risoluzione (1088px) ha permesso di rilevare con continuità anche i giocatori più distanti e piccoli.

Dal punto di vista dell'**Associazione (AssA 0.636)**, il sistema "Pure Motion" si è dimostrato robusto. Nonostante l'assenza di feature visive (ReID), la combinazione di una compensazione del movimento globale (GMC) e di una soglia di matching spaziale molto severa (`match_thresh: 0.9`) ha permesso di gestire correttamente la maggior parte delle interazioni, inclusi incroci veloci e brevi occlusioni, minimizzando gli ID switch.

4.2 Analisi Qualitativa

L'ispezione visiva dei risultati conferma la bontà delle metriche numeriche. In particolare:

- **Gestione dell'Affollamento:** Durante i calci d'angolo o le mischie in area, il tracker mantiene le identità separate grazie all'alta risoluzione di input e alla NMS calibrata (0.7), laddove configurazioni standard tendono a fondere i bounding box.

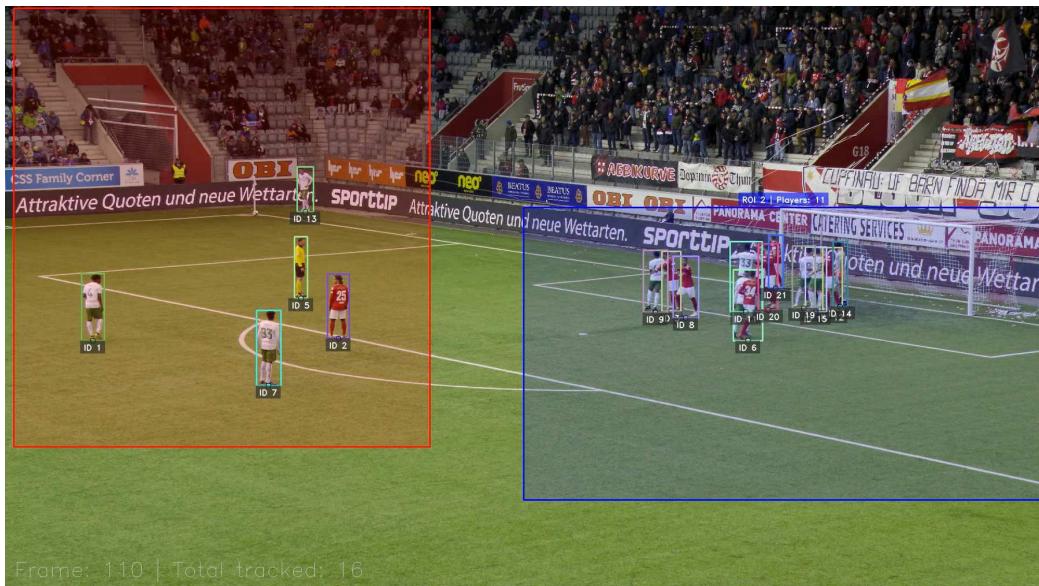


Figura 3: Esempio di gestione dell'affollamento in area di rigore. Nonostante la densità di giocatori e le parziali occlusioni, il sistema individua e traccia correttamente i singoli atleti, mantenendo distinti i bounding box grazie alla strategia "High Recall" e alla NMS ottimizzata.

- **Robustezza ai Movimenti di Camera:** Nei cambi di inquadratura repentini o zoom veloci, il modulo GMC mantiene le predizioni del tracker ancorate alla posizione reale dei giocatori, evitando la deriva delle tracce.



Figura 4: Stabilità del tracciamento durante un rapido movimento di camera. Le tracce (box colorati con ID) rimangono consistenti e ancorate ai giocatori in corsa, dimostrando l'efficacia della compensazione del movimento globale (GMC) nel prevenire ID switches causati dallo spostamento dell'inquadratura.

- **Efficacia del Field Masking:** Le visualizzazioni mostrano chiaramente come steward, fotografi e pubblico nelle prime file vengano correttamente ignorati grazie alla maschera del campo, pulendo il segnale in ingresso al tracker.



Figura 5: Efficacia del Field Masking. La maschera del campo (visibile in verde brillante nell'overlay di debug) delimita l'area di interesse. Si noti come il fotografo e lo steward a bordo campo (in basso a sinistra), pur essendo visibili, non vengano tracciati perché posizionati al di fuori della maschera segmentata, pulendo il segnale in ingresso al tracker.

5. Conclusioni e Sviluppi Futuri

5.1 Sintesi del Lavoro

In questo progetto è stata sviluppata una pipeline di Computer Vision completa per l'analisi video calcistica, affrontando le sfide della SoccerNet Benchmark Suite.

I punti chiave dell'approccio sono stati:

1. **Strategia High-Recall:** L'utilizzo di YOLO11x con soglie di confidenza minime per massimizzare il rilevamento di piccoli oggetti.
2. **Field Masking Adattivo:** Un algoritmo robusto basato su spazi colore ibridi (HSV+LAB) per filtrare i falsi positivi fuori dal campo.
3. **Tracking Pure Motion:** La dimostrazione che, in un dominio con forte omogeneità visiva (divise uguali), un tracciamento basato su geometria e movimento (BoT-SORT con GMC) può superare approcci basati su feature visive deboli.

Il risultato finale (HOA 0.742) conferma la validità di queste scelte architetturali.

5.2 Limiti e Migliorie Future

Nonostante gli ottimi risultati, il sistema presenta dei margini di miglioramento che potrebbero essere esplorati in sviluppi futuri:

- **Detector Fine-Tuning:** Sebbene YOLO11x generalizzi bene, un fine-tuning specifico sul dataset SoccerNet potrebbe migliorare ulteriormente la confidenza delle detection, permettendo di alzare le soglie di filtro senza perdere Recall e riducendo il carico computazionale del Field Masking.
- **Re-Identification Avanzato:** L'approccio attuale (Pure Motion) fallisce nelle occlusioni a lungo termine (quando un giocatore esce e rientra nell'inquadratura). L'integrazione di un modello di ReID specializzato, magari basato su **Vision Transformers (ViT)** o sul riconoscimento del **numero di maglia** (Jersey Number Recognition), permetterebbe di chiudere le tracce spezzate su lunghi intervalli temporali.
- **Deep Field Segmentation:** L'attuale maschera del campo è basata su euristiche di colore. L'adozione di una rete di segmentazione semantica (es. U-Net o DeepLabV3+) addestrata su campi da gioco garantirebbe una robustezza totale anche in condizioni di illuminazione critica o con colori del prato anomali.
- **Calibrazione della Telecamera:** Implementare una stima dell'omografia per proiettare le coordinate dei giocatori su una mappa 2D (vista dall'alto) permetterebbe di eseguire il tracking nel mondo reale invece che sul piano immagine, migliorando drasticamente la precisione del Filtro di Kalman nelle traiettorie non lineari.

Riferimenti

1. **[YOLO11]** G. Jocher e J. Qiu, *Ultralytics YOLO11*, Version 11.0.0, 2024. Disponibile su: <https://github.com/ultralytics/ultralytics>. Licenza: AGPL-3.0.
2. **[YOLOv8-FT]** EasyChamp, *EasyChamp Player Detection YOLOv8*, 2025. Hugging Face. Disponibile su: <https://huggingface.co/aabyzov/easychamp-player-detection-yolov8>.
3. **[OSNet-ReID]** B. Comandur, *Sports Re-ID: Improving Re-Identification Of Players In Broadcast Videos Of Team Sports*, arXiv preprint arXiv:2206.02373, 2022. DOI: 10.48550/ARXIV.2206.02373.