Object Tracking

Matteo Imbrosciano 1000014829

A.A. 2023/2024

Prof. Dario Allegra

Prof. Filippo Stanco

https://github.com/MatteoImbrosciano

INDICE

1 INTRODUZIONE

Nell'era digitale contemporanea, l'avvento e la proliferazione delle tecnologie video hanno catalizzato trasformazioni sostanziali in una miriade di settori, spaziando dalla sicurezza pubblica e privata ai sistemi di trasporto intelligente, dalla gestione urbana all'analisi comportamentale dettagliata. L'impiego diffuso di sistemi di videosorveglianza, droni, telecamere embedded in dispositivi mobili e altre tecnologie simili, ha reso il riconoscimento e il tracciamento automatico di oggetti non solo praticabile ma anche essenziale.

Il riconoscimento e il tracciamento di oggetti in ambienti dinamici e spesso imprevedibili presentano sfide uniche, soprattutto quando si tratta di elaborare flussi video in tempo reale. Questi flussi possono variare enormemente in termini di qualità e di complessità delle scene, rendendo cruciale l'adozione di soluzioni che coniughino precisione elevata e bassa latenza nell'elaborazione. Tali sfide amplificano la necessità di sistemi altamente efficienti e affidabili che possano garantire risultati accurati sotto pressione operativa continua.

In risposta a questi imperativi, il sistema che ho sviluppato rappresenta una convergenza di innovazioni nel campo della visione artificiale e del machine learning. Attraverso l'integrazione di algoritmi di rilevamento all'avanguardia come YOLOv5, che è noto per la sua rapidità e precisione nel riconoscere oggetti in immagini e video, il sistema assicura l'identificazione affidabile di persone e altri oggetti di interesse con una latenza minimale. Questo è particolarmente vitale in applicazioni dove la tempestività dell'identificazione può prevenire incidenti, ottimizzare il flusso del traffico o migliorare la sicurezza urbana.

Parallelamente, il nostro sistema incorpora filtri di Kalman avanzati, una metodologia consolidata per il tracciamento di oggetti in movimento che permette di predire e aggiornare la posizione di un oggetto nel tempo con grande precisione. Questo aspetto del sistema è fondamentale per mantenere il tracciamento continuo degli oggetti anche quando questi sono temporaneamente occlusi o si muovono rapidamente attraverso l'ambiente osservato. L'uso dei filtri di Kalman facilita un aggiustamento dinamico alle nuove informazioni ricevute e attenua gli errori derivanti dalle imprecisioni intrinseche delle predizioni basate esclusivamente su modelli di apprendimento automatico.

L'implementazione di queste tecnologie avanzate non solo aumenta l'efficacia del sistema in condizioni operative reali ma stabilisce anche nuovi standard per le soluzioni di tracciamento automatico, spianando la strada per future innovazioni che potrebbero ulteriormente rivoluzionare la gestione e l'analisi dei flussi video complessi. Con queste capacità, il nostro sistema si propone come uno strumento essenziale per affrontare le sfide del mondo moderno, migliorando la sicurezza, l'efficienza e la comprensione degli spazi urbani e non solo.

1.1 Obiettivo

L'obiettivo principale del nostro sistema è quello di fornire una soluzione affidabile e scalabile per il monitoraggio video che può funzionare efficacemente anche in scenari affollati e in condizioni di variabilità ambientale, come variazioni di illuminazione o di velocità degli oggetti. Utilizzando il modello YOLOv5, uno degli algoritmi di rilevamento oggetti più veloci e accurati disponibili, il sistema è in grado di identificare le auto all'interno dei frame con una precisione notevole. Per garantire un tracciamento continuo e fluido degli oggetti rilevati anche in caso di occlusioni temporanee o movimenti rapidi, abbiamo integrato un robusto sistema di filtri di Kalman. Questi filtri aiutano a prevedere e aggiornare la posizione degli oggetti, minimizzando gli errori di tracciamento e migliorando la coerenza del tracking nel tempo.

Infine, per valutare l'efficacia del nostro sistema e garantire la manutenzione della qualità visiva tra i frame processati, abbiamo implementato un insieme di metriche di valutazione video, come IoU, Precision e ReCall. Queste metriche forniscono indicatori quantitativi dell'integrità visiva e della coerenza strutturale dei video elaborati, offrendo strumenti essenziali per l'analisi di performance e per l'eventuale affinamento delle tecniche impiegate.

Attraverso questo sistema, miriamo a stabilire un nuovo standard nell'ambito del tracciamento video, fornendo agli utenti uno strumento potente per migliorare la sicurezza e l'efficacia delle operazioni di monitoraggio in tempo reale.

Il rilevamento degli oggetti è una componente fondamentale dei sistemi di visione artificiale che consente l'identificazione e la localizzazione di specifici elementi all'interno di immagini o video.

Questo processo è essenziale in molteplici applicazioni che vanno dalla sicurezza alla navigazione autonoma, dall'analisi del comportamento umano alla gestione dei contenuti multimediali.

Nel contesto del sistema che abbiamo sviluppato ovvero l'object tracking, il rilevamento degli oggetti assume un ruolo cruciale, specialmente per la sua integrazione con tecnologie di tracciamento in tempo reale e analisi video avanzata.

2.1 Tecnologie di Rilevamento

Il rilevamento degli oggetti nel nostro sistema è implementato tramite il modello YOLOv5 (You Only Look Once, versione 5), che rappresenta l'ultima evoluzione di una famosa serie di modelli di deep learning progettati per fornire un rilevamento ad alte prestazioni con tempi di elaborazione ridotti. YOLOv5 è particolarmente noto per la sua capacità di elaborare immagini in tempo quasi reale, il che lo rende ideale per applicazioni in cui la velocità è tanto cruciale quanto l'accuratezza.

2.2 Funzionamento di YOLOv5

YOLOv5 rappresenta un punto di riferimento nel campo del rilevamento oggetti grazie alla sua architettura innovativa e alla sua efficienza operativa. La peculiarità di YOLOv5, e della serie YOLO in generale, sta nella sua capacità di analizzare l'immagine complessiva in una sola passata, differenziandosi sostanzialmente da altri approcci che analizzano parti dell'immagine in fasi separate o che richiedono meccanismi di proposizione di regioni.

Il cuore dell'approccio YOLOv5 risiede nella suddivisione dell'immagine in una griglia, dove ogni cella della griglia è responsabile del rilevamento degli oggetti il cui centro cade all'interno di essa. Questa suddivisione consente al modello di mappare efficientemente lo spazio dell'immagine e di associare ogni rilevamento a una specifica area, semplificando così il compito di localizzazione.

Per ogni cella della griglia, YOLOv5 prevede simultaneamente i bounding boxes – rettangoli che delimitano l'oggetto rilevato – e le probabilità associate a ciascuna classe di oggetto che il modello è stato addestrato a riconoscere. I bounding boxes sono caratterizzati da quattro attributi: le coordinate del centro (x, y), la larghezza (w) e l'altezza (h). Ogni box ha anche associata una probabilità di classe, che indica la confidenza del modello nel rilevamento dell'oggetto all'interno di quel box.

Il nome "You Only Look Once" deriva da questa metodologia unica di valutazione complessiva dell'immagine in un unico ciclo di elaborazione, contrariamente ad altri modelli che richiedono più passaggi o che si basano su un processo iterativo di proposizione e verifica delle regioni. Questo approccio non solo accelera notevolmente il processo di rilevamento ma riduce anche la complessità computazionale, rendendo YOLOv5 particolarmente adatto per applicazioni in tempo reale e su dispositivi con capacità di elaborazione limitate.

La robustezza e l'affidabilità di YOLOv5 derivano anche dall'ampiezza e varietà dei dataset su cui viene addestrato. Dataset come COCO, PASCAL VOC e altri simili contengono milioni di immagini annotate che coprono un'ampia gamma di scenari, oggetti e contesti. L'addestramento su tali dataset consente a YOLOv5 di sviluppare una capacità eccezionale di generalizzazione, rendendolo capace di riconoscere oggetti in contesti molto diversi e spesso anche in condizioni visive complesse o sfavorevoli.

YOLOv5 rappresenta quindi non solo un modello di rilevamento oggetti di alto livello per precisione e velocità ma anche un punto di riferimento nell'evoluzione dell'apprendimento profondo e della visione artificiale. La sua capacità di combinare efficacemente velocità, precisione e affidabilità ne fa una scelta privilegiata in una varietà di applicazioni pratiche, spaziando dalla sorveglianza alla navigazione autonoma, dall'analisi video all'interazione uomo-macchina. La continua evoluzione e ottimizzazione di YOLO e dei suoi successori promettono ulteriori miglioramenti in termini di capacità di rilevamento e efficienza operativa.

2.3 La precisione e l'efficienza

La precisione e l'efficienza di YOLOv5, un modello all'avanguardia nel rilevamento degli oggetti, derivano da una serie di innovazioni tecniche e architetturali che ne ottimizzano le prestazioni mantenendo allo stesso tempo requisiti computazionali contenuti. Questi aspetti sono cruciali per applicazioni in tempo reale e su dispositivi con capacità di elaborazione limitate.

L'architettura di YOLOv5 è progettata per massimizzare l'efficienza computazionale e la precisione di rilevamento. Sfrutta convoluzioni profonde, che sono operazioni matematiche applicate alle immagini per estrarre caratteristiche visive ad alto livello. Queste caratteristiche sono fondamentali per identificare oggetti all'interno delle immagini, poiché catturano dettagli complessi che distinguono un oggetto dall'altro.

La profondità delle convoluzioni permette al modello di apprendere rappresentazioni ricche e variegate, che sono essenziali per il rilevamento accurato di oggetti in scenari diversificati.

Una delle chiavi dell'efficienza di YOLOv5 è l'impiego di AutoML (Automated Machine Learning) e pruning nelle fasi di sviluppo e ottimizzazione del modello. L'AutoML aiuta a identificare automaticamente le configurazioni ottimali della rete neurale, riducendo il tempo e le risorse necessarie per la sperimentazione manuale.

Il pruning, d'altra parte, consiste nel rimuovere i pesi meno importanti dalla rete neurale, una pratica che porta a modelli più leggeri e veloci senza un calo significativo della precisione di rilevamento.

Queste tecniche riducono la complessità del modello e i suoi requisiti di memoria, rendendo YOLOv5 particolarmente adatto per l'esecuzione su hardware con risorse limitate. Il pruning, in particolare, è cruciale per adattare il modello a dispositivi edge come telecamere di sicurezza, smartphone e sistemi embedded in veicoli autonomi, dove la rapidità di elaborazione è tanto critica quanto l'accuratezza del rilevamento.L'ottimizzazione di YOLOv5 per precisione ed efficienza si traduce in un equilibrio tra la qualità del rilevamento e la velocità di elaborazione.

Questo equilibrio è fondamentale per applicazioni in tempo reale, dove decisioni rapide possono essere cruciali. Ad esempio, in ambito di sorveglianza, un sistema in grado di identificare rapidamente un individuo sospetto può permettere un intervento tempestivo. Analogamente, nella guida autonoma, la capacità di riconoscere in tempo reale pedoni o ostacoli è vitale per la sicurezza.

In conclusione, YOLOv5 rappresenta un punto di svolta nel rilevamento degli oggetti, offrendo una combinazione ottimale di precisione, velocità ed efficienza.

Le sue innovazioni architetturali e le tecniche di ottimizzazione, come l'uso di AutoML e pruning, lo rendono adatto per un'ampia gamma di applicazioni, specialmente quelle che richiedono elaborazioni in tempo reale su dispositivi con risorse limitate. L'adozione di YOLOv5 in questi scenari promette miglioramenti significativi nella capacità di monitorare e interpretare l'ambiente circostante, aprendo nuove frontiere nell'automazione e nella sicurezza.

2.4 Integrazione con altri sistemi

L'integrazione inizia con YOLOv5 che identifica gli oggetti in un frame e ne calcola i bounding boxes. Queste informazioni sono fondamentali perché definiscono non solo la presenza dell'oggetto ma anche la sua posizione precisa all'interno dell'immagine. Una volta rilevati gli oggetti e ottenuti i loro bounding boxes, le coordinate centrali di questi boxes (calcolate come il punto medio tra i vertici superiori e inferiori) vengono fornite ai filtri di Kalman.

Ogni oggetto tracciato è associato a un filtro di Kalman distinto, che ne predice la posizione nel frame successivo. Quando YOLOv5 rileva nuovamente l'oggetto, le misurazioni aggiornate (i.e., la nuova posizione) sono utilizzate per aggiornare lo stato del filtro di Kalman, affinando così la previsione della posizione futura dell'oggetto. Questo ciclo di predizione e aggiornamento continua per tutta la durata del video, permettendo un tracciamento fluido e accurato degli oggetti in movimento.

Questa integrazione offre diversi vantaggi significativi:

- Robustezza ai Disturbi: La combinazione di YOLOv5 e filtri di Kalman rende il sistema più robusto ai disturbi e alle occlusioni temporanee, mantenendo il tracciamento anche quando l'oggetto è parzialmente nascosto o sfocato.
- Tracciamento Coerente: Il sistema garantisce un tracciamento coeso e continuo degli oggetti attraverso i frame, anche in condizioni di movimento rapido o cambiamenti improvvisi di direzione.
- Ottimizzazione delle Risorse: L'efficienza computazionale di YOLOv5, combinata con la leggerezza dei filtri di Kalman, permette l'implementazione del sistema anche su dispositivi con risorse limitate, massimizzando l'utilizzo delle capacità di elaborazione disponibili.
- Applicazioni Versatili: L'integrazione di queste tecnologie trova applicazione in una vasta gamma di scenari, dalla sicurezza alla sorveglianza, dalla navigazione autonoma all'analisi sportiva, offrendo strumenti avanzati per l'interpretazione dinamica di scene complesse.

In conclusione, l'integrazione tra YOLOv5 e i filtri di Kalman esemplifica un approccio olistico al problema del rilevamento e del tracciamento di oggetti nei video, dove l'accuratezza del rilevamento si fonde con l'intelligenza del tracciamento per creare un sistema di analisi video potente ed efficiente.

2.5 Sfide

YOLOv5 ha rappresentato un notevole passo in avanti nell'ambito del rilevamento degli oggetti mediante tecniche di visione artificiale e apprendimento profondo. Tuttavia, come avviene per ogni tecnologia all'avanguardia, ci sono ambiti di miglioramento che, se indirizzati, potrebbero ulteriormente elevare le sue capacità e applicazioni.

- **Condizioni di Illuminazione Variabile**: Una delle principali sfide è il rilevamento efficace in condizioni di scarsa illuminazione o in presenza di forti contrasti luminosi. Queste situazioni possono portare a false rilevazioni o a mancate identificazioni, poiché la variazione nell'illuminazione influisce significativamente sull'apparenza visiva degli oggetti.
- **Occlusioni e Sovrapposizioni:** Un'altra sfida critica è rappresentata dalle occlusioni e dalle sovrapposizioni di oggetti, particolarmente comuni in scene affollate. La capacità di distinguere oggetti separati quando si sovrappongono parzialmente o sono occlusi richiede un'elaborazione sofisticata e rappresenta un'area in cui YOLOv5 può essere migliorato.
- **Scalabilità e Adattabilità:** Mentre YOLOv5 eccelle nel rilevamento di oggetti per i quali è stato specificamente addestrato, la sua capacità di adattarsi a nuove classi di oggetti o a variazioni significative nell'aspetto degli oggetti noti è limitata dalla varietà dei dati su cui è stato addestrato

2.6 Miglioramenti futuri

Per affrontare queste sfide, diversi approcci di ricerca e sviluppo sono in esplorazione:

- **Illuminazione e Augmentation dei Dati:** Ampliare i set di dati di addestramento con immagini che presentano un'ampia gamma di condizioni di illuminazione o utilizzare tecniche di data augmentation che simulano tali condizioni può aiutare il modello a generalizzare meglio in ambienti diversi.
- **Rete Neurale Capsulare:** L'adozione di reti neurali capsulari, che possono mantenere informazioni gerarchiche e spaziali sugli oggetti, offre una via promettente per gestire meglio le occlusioni e le sovrapposizioni, consentendo al sistema di comprendere la composizione degli oggetti in maniera più robusta.
- **Addestramento Continuo e Apprendimento Online:** Implementare meccanismi che permettano a YOLOv5 di apprendere continuamente da nuovi dati o di adattarsi in tempo reale alle variazioni dell'ambiente potrebbe significativamente migliorare la sua versatilità e affidabilità.
- **Tecniche di Fusion Sensoriale:** Integrare dati provenienti da diversi tipi di sensori, come termici o a infrarossi, potrebbe aiutare a superare le limitazioni imposte da condizioni di illuminazione sfavorevoli, migliorando il rilevamento in un'ampia varietà di contesti ambientali.

Mentre YOLOv5 rappresenta già una solida base per il rilevamento e il tracciamento di oggetti in tempo reale, il suo continuo sviluppo e l'integrazione con tecnologie complementari apriranno nuove frontiere nelle applicazioni di visione artificiale. Attraverso la ricerca dedicata e l'innovazione tecnologica, possiamo aspettarci che le future iterazioni di YOLO e sistemi simili offrano prestazioni ancora più elevate, maggiore affidabilità e una più ampia gamma di applicabilità, spianando la strada verso soluzioni sempre più intelligenti e capaci nel campo dell'analisi video automatizzata.

2.7 Sistemi di Stabilizzazione

Un sistema di stabilizzazione è fondamentale nell'ambito della videografia e della fotografia per garantire che l'immagine o il video risultante sia il più nitido e libero da distorsioni indesiderate possibile. Questa necessità si manifesta particolarmente in contesti dinamici, come nella ripresa di scene in movimento o quando si utilizza l'equipaggiamento in condizioni instabili. I movimenti della camera, come specificato, possono essere categorizzati in "padding" e "jitter", dove il primo si riferisce a movimenti intenzionali e controllati del dispositivo di registrazione, e il secondo a movimenti non intenzionali, spesso rapidi e di breve durata, che possono causare sfocature o vibrazioni nell'immagine finale.

- **Sistemi di Stabilizzazione Analogici:** Prima dell'avvento e della diffusione dei sistemi digitali, la stabilizzazione delle immagini veniva affidata a meccanismi analogici. Questi sistemi si basano su componenti fisici come:
    - Accelerometri: Sensori che misurano l'accelerazione della camera, permettendo di compensare i movimenti bruschi.
    - Giroscopi: Dispositivi che sfruttano la rotazione per mantenere l'orientamento, fornendo un punto di riferimento stabile che aiuta a controbilanciare i movimenti involontari.
    - Sensori di Velocità Angolare: Rilevano la velocità di rotazione della camera, utili per correggere le rotazioni indesiderate.
    - Ammortizzatori Meccanici: Sistemi che fisicamente assorbono o riducono le vibrazioni, spesso utilizzati in congiunzione con i dispositivi sopra menzionati per una stabilizzazione più efficace.

  Questi componenti possono essere integrati direttamente nelle apparecchiature di registrazione o essere parte di accessori aggiuntivi, come gimbal o steadicam, che isolano la camera dai movimenti indesiderati.

- **Sistemi di Stabilizzazione Digitale (DIS):** Con l'avanzamento della tecnologia digitale, i sistemi di stabilizzazione digitale (Digital Image Stabilization, DIS) hanno guadagnato popolarità grazie alla loro versatilità e capacità di essere integrati in dispositivi più compatti, come smartphone e action camera. I DIS analizzano il contenuto video frame per frame, identificando e compensando i movimenti indesiderati attraverso l'elaborazione digitale delle immagini. Questo processo può includere:

    - **Analisi del Movimento:** Algoritmi avanzati calcolano il movimento tra i frame consecutivi, determinando la direzione e l'entità della correzione necessaria.
    - **Compensazione del Movimento:** Una volta identificato il movimento indesiderato, il sistema può stabilizzare l'immagine spostando digitalmente i frame nella direzione opposta o adattando i parametri di zoom e crop per mantenere il soggetto centrato e stabile.
    - **Rimozione del Jitter:** I DIS sono particolarmente efficaci nel ridurre il jitter, poiché possono rilevare e correggere rapidamente piccole variazioni di movimento tra i frame.

La stabilizzazione digitale offre notevoli vantaggi in termini di flessibilità e integrazione, ma può anche presentare sfide, come la perdita di qualità dell'immagine dovuta al crop o al processo di stabilizzazione digitale. Inoltre, situazioni con bassa luminosità o con movimenti molto rapidi possono rappresentare una sfida per i DIS a causa della minore precisione nell'analisi del movimento.

La ricerca e lo sviluppo nei sistemi di stabilizzazione continuano ad avanzare, con l'obiettivo di combinare l'efficacia dei sistemi analogici con la flessibilità dei sistemi digitali. Nuove tecnologie, come l'intelligenza artificiale e l'apprendimento profondo, stanno emergendo come soluzioni promettenti per migliorare ulteriormente la precisione della stabilizzazione digitale, consentendo di anticipare e compensare i movimenti in modo ancora più accurato e in scenari sempre più complessi.

2.7.1 I sistemi di stabilizzazione digitale (DIS)

I sistemi di stabilizzazione digitale (DIS) rappresentano una componente tecnologica fondamentale per migliorare la qualità visiva dei video, rendendoli meno suscettibili a vibrazioni e movimenti indesiderati che possono deteriorare l'esperienza visiva. La distinzione tra DIS real-time e DIS post-processing evidenzia la flessibilità di questi sistemi nell'adattarsi a differenti esigenze e contesti di utilizzo, dalla registrazione live all'editing video professionale. Esaminiamo più dettagliatamente le tre fasi principali su cui si basa il funzionamento dei sistemi DIS.

Fase 1: Stima del Movimento

Il processo inizia con l'analisi dettagliata dei frame per identificare il movimento all'interno del video. Questa fase è cruciale perché permette di comprendere come gli oggetti e l'ambiente cambiano da un frame all'altro. La stima del movimento si avvale di algoritmi che tracciano i cambiamenti di posizione degli oggetti, sfruttando i pixel o le caratteristiche salienti delle immagini per calcolare i vettori di movimento. Questo passaggio è fondamentale per determinare non solo la direzione e la velocità del movimento ma anche per distinguere tra i movimenti intenzionali (padding) e quelli non intenzionali (jitter).

Fase 2: Filtraggio del Movimento

Dopo aver stimato i motion vector locali (LMV) affidabili, il passo successivo consiste nel calcolare il motion vector globale (GMV) della scena. Il GMV rappresenta la direzione e l'intensità complessiva del movimento che si desidera correggere o mantenere, a seconda che sia classificato come jitter o come padding. L'approccio alla stabilizzazione diventa relativamente semplice se l'obiettivo è mantenere la camera immobile: attraverso l'applicazione di trasformazioni inverse basate sui GMV, è possibile riportare punti specifici della scena sempre nella stessa posizione frame dopo frame,

neutralizzando così i movimenti indesiderati. La capacità di distinguere tra padding e jitter diventa fondamentale qui: mentre i movimenti intenzionali possono essere mantenuti, quelli non intenzionali vengono corretti per stabilizzare il video.

Filtro di Kalman

Il filtro di Kalman ha le seguenti caratteristiche:

- Prende in input una serie di misure osservate nel tempo
- Tiene conto di un eventuale rumore casuale
- Predizione: fissato un tempo stima la misura
- Correzione: osservata una misura può correggerla, eventualmente basandosi su una predizione

La versione discreta di questo filtro si basa su due equazioni differenziali stocastiche in cui lo stato x_k è dato da:

$$X_t = Ax_{t-1} + Bu_t + w_t$$

La misura z_k di x_k è data da:

$$Z_t = Hx + v$$

Dove t rappresenta l'istante temporale; A,B,H sono modelli transazionali; w,v rappresentano il rumore gaussiano e u rappresenta il controllo dell'utente. La predizione e la correzione si ottengono affiancando a questo filtro un tracciatore bayesiano che permette di stimare le probabilità.

Fase 3: Deformazione (Post-processing) dell'Immagine

L'ultima fase del processo di stabilizzazione digitale è la deformazione o post-processing dell'immagine. Questa fase si occupa di adattare il frame per compensare le correzioni applicate durante il filtraggio del movimento. Può includere operazioni come il cropping, il ridimensionamento o la rotazione dei frame per mantenere la continuità visiva e assicurare che il video finale appaia fluido e privo di discontinuità causate dalla stabilizzazione. Il post-processing è essenziale per garantire che l'esperienza visiva finale sia piacevole e libera da distorsioni percepite come artificiali o disturbanti.

Le tre fasi descritte formano un processo complesso che richiede un'attenta gestione dei dati e un'elaborazione sofisticata per ottenere risultati ottimali. Nel contesto del tuo progetto, l'efficacia della stabilizzazione dipenderà significativamente dalla precisione della stima del movimento e dalla capacità del sistema di filtrare in modo selettivo i movimenti, preservando l'intenzionalità del cameraman mentre si eliminano le vibrazioni e i movimenti indesiderati. L'avanzamento nelle tecnologie di visione artificiale e l'incremento della potenza computazionale disponibile aprono nuove possibilità per rendere i DIS ancora più precisi ed efficaci, potenziando le applicazioni pratiche in numerosi campi, dalla produzione cinematografica all'uso quotidiano in dispositivi mobili.

2.8 Metriche di Qualità

Le metriche di qualità per le immagini e i video giocano un ruolo fondamentale nell'ambito della visione artificiale e del processing multimediale. Forniscono un ponte tra l'aspetto tecnico e quantitativo dell'elaborazione delle immagini e la percezione umana della qualità visiva. Eseguire una valutazione qualitativa efficace richiede una comprensione profonda sia degli aspetti tecnici che di quelli percettivi.

Bisogna distinguere due diverse tipologie di qualità:

- **Qualità assoluta**: La qualità assoluta di un'immagine si riferisce alle sue proprietà oggettive, misurabili attraverso un insieme di attributi che possono variare da quelli semplici, come il contrasto o la luminosità, a quelli più complessi composti da sottocategorie. Alcuni di questi attributi sono direttamente quantificabili, mentre altri, per la loro natura intrinseca, restano al di fuori dell'ambito della misurazione diretta.
- **Qualità Percepita**: la qualità percepita incorpora la dimensione soggettiva della visione, il che significa che oltre alle caratteristiche fisiche dell'immagine, intervengono fattori come l'interpretazione personale, le aspettative e la familiarità dell'osservatore con il soggetto dell'immagine. Questo tipo di qualità è influenzato da vari elementi, tra cui le tecniche di elaborazione utilizzate, le specifiche tecnologiche del dispositivo di visualizzazione e le caratteristiche uniche del sistema visivo umano.

Dunque, le metriche di qualità assoluta vengono definite metriche oggettive, mentre quelle di qualità percepita vengono definite metriche soggettive.

Le metriche di qualità possono essere a sua volta classificate in base alla presenza di una maggiore o minore disponibilità del segnale di riferimento, con il quale confrontare l'immagine distorta.

A seconda della disponibilità dell'immagine di riferimento, le metriche di qualità si dividono in:

- **Full Reference (FR):** Queste metriche presuppongono la disponibilità dell'immagine originale non distorta, consentendo un confronto diretto e completo. Offrono una valutazione precisa della qualità, confrontando l'immagine distorta con quella di riferimento su una base pixel per pixel o attraverso caratteristiche estratte.

- **No-Reference (NR):** Tali metriche sono cruciali quando l'immagine di riferimento non è disponibile. Utilizzano modelli predittivi o estraggono indicatori di qualità basandosi unicamente sull'immagine distorta, facendo affidamento su principi di qualità intrinseci o su assunzioni sulle caratteristiche delle immagini di alta qualità.

- **Reduced Reference (RR):** Questo approccio si colloca tra i due precedenti, utilizzando solo informazioni parziali estratte dall'immagine originale. Le metriche RR sono particolarmente utili in scenari in cui si desidera bilanciare la necessità di un riferimento con la limitata disponibilità di dati o con restrizioni di banda.

2.8.1 IoU (Intersection over Union)

La metrica IoU (Intersection over Union), anche conosciuta come Jaccard Index, è una metrica comunemente utilizzata per valutare la qualità delle previsioni di modelli di segmentazione o rilevamento oggetti in computer vision.

L'IoU viene calcolata come il rapporto tra l'area dell'intersezione e l'area dell'unione tra due regioni. Formalmente, l'IoU tra due regioni A e B è calcolata come:

$$IoU(A, B) = \frac{A \cap B}{A \cup B}$$

Dove:

• $A \cap B$ rappresenta l'area dell'intersezione tra le due regioni.

• $A \cup B$ rappresenta l'area dell'unione tra le due regioni.

Nel contesto della segmentazione o del rilevamento oggetti, l'IoU viene spesso utilizzata come metrica di valutazione per misurare quanto bene le previsioni del modello corrispondano alle ground truth (ovvero le verità di terreno) annotate dagli esseri umani.

2.8.2 Accuracy

L'accuracy, o accuratezza, è una delle metriche più comuni per valutare le prestazioni di un modello di classificazione, ma può essere applicata anche in contesti come il rilevamento di oggetti.
L'accuracy misura la frazione di predizioni corrette (sia positive che negative) rispetto al totale delle predizioni fatte dal modello.

$$\text{Accuracy} = \frac{\text{TP}+\text{TN}}{\text{TP}+\text{TN}+\text{FP}+\text{FN}}$$

Dove:

- TP (True Positives) sono i casi in cui il modello predice correttamente la presenza della classe positiva;
- TN (True Negatives) sono i casi in cui il modello predice correttamente l'assenza della classe positiva;
- FP (False Positives) sono i casi in cui il modello predice erroneamente la presenza della classe positiva;
- FN (False Negatives) sono i casi in cui il modello predice erroneamente l'assenza della classe positiva.
-

Nel rilevamento di oggetti, l'uso dell'accuracy può essere meno diretto perché non esiste sempre un concetto chiaro di "negativo vero" (TN). In molti sistemi di rilevamento di oggetti, ogni area dell'immagine potenzialmente contiene un oggetto di interesse, quindi il concetto di TN (aree correttamente identificate come prive di oggetti) spesso non è applicabile. Di conseguenza, altre metriche come la Precision e il Recall diventano più rilevanti:

- Precision: indica quanto sono affidabili i positivi predetti dal modello.
- Recall: indica quanto bene il modello è in grado di identificare tutti i positivi rilevanti.

2.8.3 Precision

La precisione è una delle metriche più comuni utilizzate per valutare le prestazioni di un modello di classificazione. Rappresenta la frazione di istanze classificate correttamente come positivi rispetto a tutte le istanze classificate come positivi, indipendentemente dalla classificazione negativa.

$$\text{Precisione} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

Dove:

• TP (True Positives) rappresenta il numero di casi positivi che sono stati correttamente classificati come positivi dal modello.

• FP (False Positives) rappresenta il numero di casi negativi che sono stati erroneamente classificati come positivi dal modello.

La precisione misura la proporzione di previsioni positive del modello che sono corrette.

Una precisione più alta indica che il modello ha una bassa percentuale di falsi positivi, ossia ha meno casi negativi erroneamente classificati come positivi.

Una precisione più bassa indica che il modello ha una percentuale più alta di falsi positivi.

2.8.4 ReCall

La recall, conosciuta anche come sensitivity o true positive rate (TPR), è una metrica utilizzata per valutare le prestazioni di un modello di classificazione, specialmente in presenza di classi sbilanciate. Rappresenta la frazione di istanze positive correttamente identificate dal modello rispetto a tutte le istanze positive effettive.

In formule si ha:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

Dove:

- TP (True Positives) rappresenta il numero di casi positivi che sono stati correttamente classificati come positivi dal modello.
- FN (False Negatives) rappresenta il numero di casi positivi che sono stati erroneamente classificati come negativi dal modello.

La recall misura la capacità del modello di identificare correttamente tutti i casi positivi presenti nel dataset.

Una recall più alta indica che il modello ha una bassa percentuale di casi positivi non individuati (falsi negativi).

Una recall più bassa indica che il modello ha una percentuale più alta di casi positivi non individuati.

2.9 Considerazioni Finale

Il sistema proposto sfrutta le potenzialità offerte dalle più recenti innovazioni nel campo della visione artificiale e dell'intelligenza artificiale, rappresentando un'efficace risposta alle crescenti esigenze di monitoraggio video in contesti dinamici e spesso imprevedibili. L'impiego di algoritmi avanzati come YOLOv5, abbinato all'uso di filtri di Kalman per un preciso tracciamento degli oggetti in movimento, evidenzia un approccio integrato che ottimizza le prestazioni in termini di velocità, accuratezza e affidabilità.

La capacità di elaborare in tempo reale flussi video di alta qualità ha implicazioni significative in molti ambiti, dalla sicurezza pubblica e privata all'ottimizzazione dei flussi di traffico, dall'analisi comportamentale in contesti commerciali e sociali alla gestione e al coordinamento in situazioni di emergenza. La precisione nell'identificazione e nel tracciamento di persone o oggetti specifici può prevenire incidenti, facilitare indagini, migliorare l'efficienza operativa e aumentare la sicurezza generale.

La configurazione del sistema, che stabilisce nuovi standard di performance, apre la strada a un'evoluzione continua nel modo in cui interpretiamo e interagiamo con i dati video. Questo progresso non è limitato solamente all'incremento delle capacità tecniche ma include anche l'ampliamento delle possibilità applicative dei sistemi di monitoraggio e analisi video.

L'avanzamento tecnologico è un processo in continuo movimento, e il campo della visione artificiale non fa eccezione. L'introduzione di algoritmi di apprendimento profondo sempre più sofisticati, l'elaborazione edge per ridurre la latenza nelle decisioni in tempo reale, e l'integrazione di sensoristica avanzata (come i sensori termici o multispettrali) possono ulteriormente migliorare le capacità del sistema. Inoltre, l'adozione di infrastrutture computazionali distribuite e l'utilizzo di tecnologie blockchain per la sicurezza dei dati sono solo alcuni degli sviluppi potenziali che potrebbero elevare le prestazioni e l'affidabilità dei sistemi di monitoraggio video.

Implementare sistemi avanzati di monitoraggio video ha un impatto profondo sulla sicurezza e sulla gestione degli ambienti urbani e oltre. La capacità di analizzare in tempo reale vasti flussi di dati video consente di rispondere rapidamente a eventi critici, migliorare la pianificazione urbana e le strategie di intervento in caso di emergenze, nonché di ottimizzare i servizi e le infrastrutture cittadine. Questo non solo accresce la sicurezza pubblica ma contribuisce anche a un'esperienza urbana più sicura, efficiente e piacevole per i cittadini.

L'adozione e l'integrazione di tecnologie avanzate nel monitoraggio video stanno definendo un nuovo paradigma nell'analisi e nella gestione dei dati visivi. Il sistema che abbiamo sviluppato non solo affronta le sfide attuali ma pone le basi per future innovazioni, dimostrando come il progresso tecnologico possa essere diretto verso soluzioni che migliorano la sicurezza, l'efficienza e la qualità della vita nelle nostre società. L'evoluzione futura in questo campo promette di aprire nuovi orizzonti applicativi, spingendoci verso una comprensione sempre più integrata e dinamica del mondo che ci circonda.

3 Applicazioni pratiche

Il progetto si concentra sull'analisi dettagliata e sul filtraggio del movimento nei video, con l'obiettivo di esplorare e valutare le capacità degli algoritmi di visione artificiale e di trattamento delle immagini in contesti realistici. Attraverso l'applicazione di queste tecnologie a campioni di video che riproducono scenari reali, il progetto mira a offrire una panoramica completa dell'efficacia delle soluzioni adottate, considerando vari fattori che influenzano le prestazioni e l'applicabilità pratica delle tecniche di filtraggio del movimento.

L'analisi dettagliata e il filtraggio del movimento nei video attraverso questo approccio metodico non solo evidenziano l'efficacia delle tecnologie di visione artificiale ma aprono anche la strada a ulteriori ricerche e sviluppi. L'obiettivo è quello di perfezionare continuamente le tecniche di analisi video per affrontare le sfide emergenti e sfruttare pienamente il potenziale di queste tecnologie avanzate in applicazioni pratiche, migliorando così la sicurezza, l'efficienza e l'efficacia del monitoraggio video in una vasta gamma di contesti operativi. Quindi cercherò di valutare le mie prestazioni in base al rilevamento delle mie Bounding Box predette con quelle reali. Ho utilizzato un dataset mo_labels.csv in cui per ogni frame avrò l'etichetta corretta delle coordinate del mio Bounding Box che andrò a confrontare con quelle che andrò a predirre.

3.1 Fasi del Progetto

Al fine di garantire un'analisi esaustiva, il lavoro è stato strutturato in diverse fasi per consentire un'esplorazione approfondita di ciascuna di esse. Questa decisione è stata presa per semplificare il processo e permettere una valutazione completa e dettagliata di ogni aspetto coinvolto.

3.1.1 Caricamento e pulizia del mio dataset

Vado a leggere il mio dataset.

```
mot_labels = pd.read_csv('C:\\Users\\matte\\OneDrive\\Desktop\\Multimedia\\progetto\\mot_labels.csv')
```

Vado a filtrare le mie informazioni relative alle 'car' in base al video che prendo in considerazione:

- **Video_1:**

```
# Pulire i dati rimuovendo le righe con valori mancanti e filtrando per 'car'
cleaned_mot_labels = mot_labels.dropna()
df_real_labels = cleaned_mot_labels[
    (cleaned_mot_labels['videoName'] == '00c4c672-26d36ad8') &
    (cleaned_mot_labels['category'] == 'car')
]
```

Successivamente, vado a salvarmi in 'label_reali_1" le coordinate del mio bounding Box per ogni frame, che mi serviranno successivamente per confrontare con quelle predette e andarle a valutare con delle metriche opportune.

```
# Selezione delle colonne pertinenti
df_real_labels = df_real_labels[['frameIndex', 'box2d.x1', 'box2d.x2', 'box2d.y1', 'box2d.y2']]

# Salvataggio in un nuovo CSV
output_path = 'labels_reali_1.csv'
```

- **Video_2:**

```
# Pulire i dati rimuovendo le righe con valori mancanti e filtrando per 'car'
cleaned_mot_labels = mot_labels.dropna()
df_real_labels = cleaned_mot_labels[
    (cleaned_mot_labels['videoName'] == '00c17a92-d4803287') &
    (cleaned_mot_labels['category'] == 'car')
]
```

Successivamente, vado a salvarmi in 'label_reali_2.csv" le coordinate del mio bounding Box per ogni frame, che mi serviranno successivamente per confrontare con quelle predette e andarle a valutare con delle metriche opportune.

```
# Selezione delle colonne pertinenti
df_real_labels = df_real_labels[['frameIndex', 'box2d.x1', 'box2d.x2', 'box2d.y1', 'box2d.y2']]

# Salvataggio in un nuovo CSV
output_path = 'labels_reali_2.csv'
```

- **Video_3:**

```
# Rimuovere righe con valori mancanti
mot_labels.dropna(inplace=True)

# Verificare il tipo di dati aggiornato e le righe rimanenti
print(mot_labels.info())
print(mot_labels.head())


# Pulire i dati rimuovendo le righe con valori mancanti e filtrando per 'car'
cleaned_mot_labels = mot_labels.dropna()
df_real_labels = cleaned_mot_labels[
    (cleaned_mot_labels['videoName'] == '00c29c52-f9524f1e') &
    (cleaned_mot_labels['category'] == 'car')
]
```

Successivamente, vado a salvarmi in 'label_reali_3.csv" le coordinate del mio bounding Box per ogni frame, che mi serviranno successivamente per confrontare con quelle predette e andarle a valutare con delle metriche opportune.

```
# Selezione delle colonne pertinenti
df_real_labels = df_real_labels[['frameIndex', 'box2d.x1', 'box2d.x2', 'box2d.y1', 'box2d.y2']]

# Salvataggio in un nuovo CSV
output_path = 'labels_reali_3.csv'
```

3.1.2 Rilevamento degli oggetti con YOLO

Il primo passo nel processo di stima del movimento consiste nel rilevare gli oggetti presenti nei frame video. A questo scopo, il sistema impiega il modello YOLOv5, un algoritmo di deep learning noto per la sua velocità e precisione nel riconoscere oggetti in immagini e video. YOLOv5 analizza l'immagine complessiva in una singola passata, predice i bounding boxes (contenitori rettangolari) e le probabilità di classe per ogni oggetto rilevato.

Per ottimizzare ulteriormente la performance del modello in scenari reali, dove la qualità dell'immagine può variare significativamente, viene applicato un filtro mediano prima del passaggio di rilevamento. Questo filtro riduce il rumore di fondo e migliora il contrasto tra gli oggetti e il loro ambiente, facilitando la loro identificazione da parte di YOLOv5.

3.1.2 Tracciamento degli oggetti con il filtro Kalman

Una volta rilevati gli oggetti, il sistema procede al loro tracciamento utilizzando il filtro di Kalman. Questo approccio matematico fornisce una stima ottimale dello stato di un sistema dinamico in presenza di incertezze, basandosi su una serie di misurazioni osservate nel tempo, che possono essere soggette a rumore. Nel contesto del nostro sistema, ogni oggetto rilevato viene associato a un filtro di Kalman, che stima la sua posizione e velocità in ogni frame successivo.

Il filtro di Kalman opera in due fasi principali: predizione e aggiornamento. Nella fase di predizione, il filtro stima la posizione futura dell'oggetto basandosi sul suo stato attuale e sulle leggi del moto. Successivamente, quando viene effettuata una nuova rilevazione dell'oggetto, la fase di aggiornamento utilizza queste nuove informazioni per correggere la stima e ridurre l'errore.

3.1.3 Stima del Movimento

La combinazione di YOLOv5 per il rilevamento e del filtro di Kalman per il tracciamento consente una stima efficace del movimento degli oggetti nei video. Il sistema è in grado di monitorare gli spostamenti degli oggetti da un frame all'altro, determinando sia le traiettorie che le variazioni di velocità. Questa capacità di stima del movimento è fondamentale per applicazioni che richiedono una comprensione dettagliata della dinamica degli oggetti, come la sorveglianza o l'analisi del comportamento. Il sistema proposto implementa un approccio ibrido per il rilevamento e il tracciamento di oggetti in video, combinando il modello di deep learning YOLOv5 per il rilevamento oggetti e i filtri di Kalman per il loro tracciamento nel tempo. La scelta di YOLOv5 deriva dalla sua efficienza e precisione nel riconoscere oggetti in tempo reale, mentre l'utilizzo dei filtri di Kalman offre un metodo affidabile per stimare la traiettoria degli oggetti tra i frame, fornendo una stima continua del loro stato.

Componenti del Sistema:

- **ObjectTracker:** Classe principale che gestisce il tracciamento degli oggetti identificati nel video. Mantieni traccia dei punti centrali degli oggetti, degli identificativi unici, dei filtri di Kalman associati a ciascun oggetto e del conteggio dei frame persi.

- **Filtro di Kalman:** Utilizzato per prevedere e aggiornare lo stato degli oggetti tracciati. Ogni oggetto viene associato a un filtro di Kalman che ne stima la posizione (x, y) e la velocità (dx, dy) in base alle osservazioni precedenti e alle nuove misurazioni. Il filtro lavora in due fasi: predizione e correzione, permettendo di mantenere il tracciamento anche in condizioni di parziale occlusione o di movimento rapido dell'oggetto.

Rilevamento degli Oggetti: All'inizio, il video viene analizzato frame per frame. Ogni frame viene pre-processato con un filtro mediano per ridurre il rumore e migliorare la qualità del rilevamento. Successivamente, viene utilizzato YOLOv5 per identificare gli oggetti di interesse, come persone o veicoli, e per calcolare i loro bounding boxes.

Inizializzazione e Aggiornamento dei Filtri di Kalman: Per ogni oggetto rilevato, si verifica se può essere associato a un filtro di Kalman esistente (e quindi a un oggetto già tracciato nei frame precedenti) basandosi sulla distanza minima tra i punti centrali. Se un oggetto non può essere associato a nessun filtro esistente, viene inizializzato un nuovo filtro di Kalman. I filtri vengono poi aggiornati con le nuove posizioni rilevate.

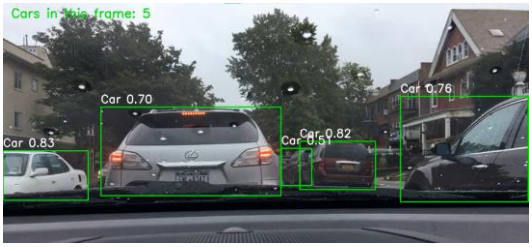
Pulizia e Gestione degli Oggetti: Dopo ogni aggiornamento, il sistema verifica quali oggetti sono stati persi (ovvero non rilevati nei frame successivi). Se un oggetto non viene rilevato per un numero di frame superiore a una soglia predefinita, il relativo filtro di Kalman viene eliminato, e l'oggetto viene rimosso dal tracciamento.

Otterrò delle coordinate del Bounding Box, frame per frame e salverò questi valori in un file “tracked_result_1.csv” che successivamente mi serviranno per la valutazione. Farò lo stesso processo per i tre video che ho analizzato, ottenendo tre file diversi: “tracked_result_1.csv”, “tracked_result_2.csv” e “tracked_result_3.csv”.

Elaboro anche un “tracked_video” per ognuno dei tre video, in cui si vedranno gli oggetti che andrò a rilevare e a tracciare.

- **Video_1:**

- **Video_2:**



- **Video_3:**

3.1.4 Valutazione tramite metriche di qualità

Nel tuo progetto di rilevamento di oggetti, l'utilizzo di metriche come l'IOU (Intersection Over Union), l'accuracy, la precisione e il recall è fondamentale per valutare compiutamente l'efficacia del modello di riconoscimento. Ogni metrica offre una prospettiva diversa sulle prestazioni e insieme forniscono un quadro completo di come il modello si comporta in diversi aspetti della rilevazione. Noi cercheremo di rilevare l'oggetto, salvare le coordinate del bounding box e confrontarle con quelle reali per ogni frame.

```
df_real_labels = pd.read_csv(r'C:\Users\matte\OneDrive\Desktop\Multimedia\prog\labels_reali_1.csv')
tracked_results = pd.read_csv(r'C:\Users\matte\OneDrive\Desktop\Multimedia\prog\tracked_results_1.csv')
```

Per ogni video io calcolo le seguenti metriche:

```
precision = true_positives / (true_positives + false_positives)
recall = true_positives / (true_positives + false_negatives)
mean_iou = sum(iou_scores) / len(iou_scores)
accuracy = true_positives / (true_positives + false_positives + false_negatives)
```

Esaminiamo il ruolo e l'importanza di ciascuna metrica:

1. Intersection Over Union (IOU)

L'IOU è particolarmente cruciale nel rilevamento di oggetti perché misura quanto bene una bounding box predetta si sovrappone con la verità di ground (la bounding box reale). Un IOU elevato indica che la posizione e la dimensione della bounding box predetta sono molto accurate. L'IOU è usato per determinare se una previsione è un vero positivo o un falso positivo, spesso con una soglia (come 0.5) per decidere questo aspetto. Questa metrica è diretta e molto intuitiva per il rilevamento di oggetti, poiché valuta l'accuratezza geometrica delle predizioni.

2. Accuracy

Sebbene l'accuracy non sia sempre diretta da calcolare nel rilevamento di oggetti a causa dell'assenza di veri negativi chiari, essa fornisce comunque un'indicazione utile sull'efficienza generale del modello nel fare previsioni corrette (vero positivo) rispetto al totale delle predizioni. Questa metrica è utile quando si desidera una misura rapida delle prestazioni complessive del modello, ma deve essere considerata con cautela e in combinazione con altre metriche per non trarre conclusioni fuorvianti, soprattutto in presenza di un dataset sbilanciato.

3. Precision

La precision è essenziale quando è importante minimizzare i falsi positivi. Ad esempio, in applicazioni di sicurezza o di sorveglianza, un alto tasso di falsi positivi potrebbe portare a allarmi inutili e spese gestionali eccessive. La precision indica la proporzione di identificazioni corrette (veri positivi) tra tutte le identificazioni positive fatte dal modello (sia veri che falsi positivi). Questa metrica aiuta a capire quanto sia affidabile il modello quando afferma di aver trovato un oggetto.

4. Recall

Il recall è critico quando è essenziale identificare tutti i casi positivi; è importante in scenari come la rilevazione medica di tumori o altre applicazioni dove mancare un oggetto potrebbe avere gravi conseguenze. Il recall misura la proporzione di veri positivi rilevati rispetto al numero totale di casi positivi effettivi (veri positivi più falsi negativi). Questa metrica valuta l'abilità del modello di trovare tutti gli oggetti rilevanti nell'immagine.

4 Risultati

- Il valore **dell'IoU** varia da 0 a 1, dove 0 indica nessuna sovrapposizione e 1 indica una sovrapposizione perfetta tra le bounding boxes.Valori tipici per un buon risultato:
    - Sotto lo 0.5: Generalmente considerato insufficiente per la maggior parte delle applicazioni.
    - 0.5 - 0.7: Accettabile, mostra una sovrapposizione moderata ma può essere migliorato.
    - Oltre 0.7: Buono, indica che la bounding box predetta copre l'oggetto molto bene rispetto alla verità di ground.
    - Oltre 0.9: Eccellente, quasi perfetta sovrapposizione delle bounding boxes.

- **Accuracy** varia da 0 a 1, dove 0 indica che tutte le predizioni sono errate, e 1 indica che tutte le predizioni sono corrette.
    - Valori tipici per un buon risultato:
    - Sotto lo 0.6: Generalmente considerato basso per la maggior parte delle applicazioni.
    - 0.6 - 0.75: Moderato, indica una precisione accettabile ma migliorabile.
    - Oltre 0.75: Buono, mostra che il modello ha una buona capacità di fare predizioni corrette.
    - Oltre 0.9: Eccellente, indica un'alta precisione nelle predizioni del modello.

- **Precision** è una metrica essenziale per valutare la qualità delle previsioni di un modello di classificazione o rilevamento oggetti. Il valore della precision varia da 0 a 1, dove 0 indica che tutte le predizioni positive sono incorrette, e 1 indica che tutte le predizioni positive sono corrette. Valori tipici per un buon risultato:
    - Sotto lo 0.5: Considerato basso per la maggior parte delle applicazioni.
    - 0.5 - 0.7: Moderato, indica una certa capacità del modello di identificare correttamente le classi positive, ma c'è spazio per miglioramenti.
    - Oltre 0.7: Buono, mostra che il modello è abbastanza affidabile nel predire le classi positive.
    - Oltre 0.9: Eccellente, indica un'elevata affidabilità nelle predizioni positive del modello.

- **ReCall** , conosciuta anche come sensibilità o tasso di vero positivo, è una metrica fondamentale nella valutazione delle prestazioni di modelli di classificazione o rilevamento oggetti. Il valore della recall varia da 0 a 1, dove 0 indica che nessun positivo reale è stato identificato correttamente, e 1 indica che tutti i positivi reali sono stati identificati.
    - Valori tipici per un buon risultato:
    - Sotto lo 0.5: Considerato basso, indica che molti casi positivi non vengono catturati dal modello.
    - 0.5 - 0.7: Moderato, mostra una capacità ragionevole del modello di identificare i casi positivi, ma con ampio margine di miglioramento.
    - Oltre 0.7: Buono, suggerisce che il modello è abbastanza efficace nel rilevare i casi positivi.
    - Oltre 0.9: Eccellente, indica una grande capacità del modello di rilevare quasi tutti i casi positivi.

- **Video_1:**

```
Accuracy: 0.7109
Precision: 0.8318
Recall: 0.8303
IOU medio: 0.8512
```

- **Video_2:**

```
Accuracy: 0.7488
Precision: 0.8581
Recall: 0.8545
IOU medio: 0.8722
```

- **Video_3:**

```
Accuracy: 0.6749
Precision: 0.8190
Recall: 0.7932
IOU medio: 0.8233
```

5 Conclusione

Possiamo vedere le prestazioni di un sistema di rilevamento oggetti valutate su tre diversi video, attraverso quattro metriche: accuracy, precision, recall e IOU medio. Ecco una conclusione basata sui dati:

- Video 1:
    - Accuracy: 0.7109 - Questa è la più bassa tra i tre video, indicando che il sistema ha avuto il maggior numero di errori totali (falsi positivi e falsi negativi) su questo video. Ciò può essere dovuto a diverse difficoltà intrinseche nel video, come sfondi complessi, occlusione degli oggetti o movimento rapido.
    - Precision: 0.8318 - Abbastanza alta, mostra che la maggior parte delle identificazioni positive è corretta. Ciò suggerisce che il sistema è affidabile quando identifica un oggetto come presente.
    - Recall: 0.8303 - Anch'essa alta, quasi paritaria alla precisione, suggerisce che il sistema è stato capace di rilevare la maggior parte degli oggetti reali. Questo è positivo, soprattutto in contesti dove è critico non perdere oggetti positivi.
    - IOU medio: 0.8512 - Questo valore molto elevato implica che le bounding boxes predette dal sistema corrispondono strettamente a quelle reali, indicando una precisione geometrica eccellente.
- Video 2:
    - Accuracy: 0.7488 - La più alta tra i tre, suggerisce che il sistema ha avuto un bilancio complessivamente migliore tra tutti i tipi di predizioni corrette e errate su questo video.
    - Precision: 0.8581 - Questa è la più alta precisione tra i video, implicando un'eccellente affidabilità nelle identificazioni positive e una minima quantità di falsi positivi.
    - Recall: 0.8545 - Simile alla precisione e relativamente alta, mostra che il sistema è riuscito a rilevare un'ampia maggioranza degli oggetti reali presenti.
    - IOU medio: 0.8722 - Il più alto dei tre, indicando che le posizioni delle bounding boxes sono state molto accurate e che vi è stata un'ottima sovrapposizione con le verità di ground.

- **Video 3:**
  - Accuracy: 0.6749 - Indica che il sistema ha avuto una percentuale più alta di errori complessivi su questo video. Questo potrebbe suggerire che il video presenta sfide più complicate, come variabilità nell'illuminazione, nell'angolazione di ripresa o nella densità di oggetti presenti.

  - Precision: 0.8190 - Nonostante l'accuracy più bassa, la precisione rimane alta, suggerendo che quando il sistema rileva un oggetto, è piuttosto probabile che sia corretto. Tuttavia, ciò non esclude la possibilità che il sistema stia perdendo alcuni oggetti reali.

  - Recall: 0.7932 - Più basso rispetto agli altri video, indicando che il sistema ha mancato una percentuale maggiore di oggetti reali. Questo potrebbe essere un'area per il miglioramento, forse affinando il modello o adattando i parametri di rilevamento.

  - IOU medio: 0.8233 - Anche se è il più basso, rimane un valore alto, il che significa che la qualità del posizionamento delle bounding boxes è comunque buona, sebbene leggermente inferiore rispetto agli altri video.

Il sistema mostra una tendenza a mantenere una precisione piuttosto elevata attraverso i diversi video, il che è indicativo di un'ottima capacità di evitare falsi positivi. La variazione nell'accuracy e nella recall suggerisce che il sistema potrebbe avere difficoltà in determinate condizioni o scenari presenti nei diversi video.

Per esempio, il Video 3 potrebbe presentare sfide che impattano negativamente sul rilevamento degli oggetti, portando a un maggior numero di falsi negativi o a una minore accuracy generale.

Il Video 2 appare come il caso di successo più marcato, con prestazioni superiori in tutte le metriche. Potrebbe essere interessante analizzare le caratteristiche di questo video che hanno contribuito a tali risultati. Forse gli oggetti erano più grandi o meno occlusi, o forse le condizioni di illuminazione erano più favorevoli per il rilevamento.

Per migliorare ulteriormente il sistema, un'analisi approfondita degli errori specifici commessi nei tre video potrebbe fornire spunti per l'ottimizzazione. Ad esempio, l'addestramento del modello su più esempi simili a quelli presenti nel Video 3 potrebbe migliorare l'accuracy e il recall in scenari simili. Inoltre, tecniche di data augmentation o l'uso di un insieme di dati più variegato potrebbero aumentare la robustezza del sistema alle variazioni tra diversi video.