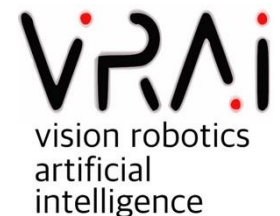


---

# Use case: Intelligenza artificiale a supporto della Dwell Time Analysis



UNIVERSITÀ  
POLITECNICA  
DELLE MARCHE



# Use Case: Dwell Time Analysis for People Flow



# Use Case: Dwell Time Analysis for People Flow

## Introduzione del progetto

### Descrizione del progetto

- Il progetto "Dwell Time Analysis for People Flow", parte dell'iniziativa CTE SQUARE Pesaro, mira a ottimizzare gli spazi pubblici utilizzando tecnologie avanzate di Computer Vision e Deep Learning. L'obiettivo principale è monitorare e analizzare i flussi di persone per ridurre i tempi di attesa e comprendere meglio i comportamenti umani, migliorando così l'efficienza e l'utilizzo delle aree pubbliche.

### Obiettivi

#### Monitorare e analizzare i flussi di persone

- Per identificare i punti critici e migliorare l'efficienza degli spazi.

#### Ridurre i tempi di attesa

- Comprendendo i comportamenti e ottimizzando l'uso degli spazi.

#### Ottimizzare l'utilizzo degli spazi pubblici

- Utilizzando soluzioni tecnologiche avanzate.

### Principali aree di intervento

#### Analisi dei Tempi di Attesa

- Monitorare il tempo che le persone trascorrono in specifiche aree per identificare i punti critici.

#### Ottimizzazione dei Modelli di Deep Learning per Dispositivi Edge

- Adattare i modelli di deep learning per funzionare efficacemente su dispositivi a bassa potenza.

#### Implementazione Pratica

- Applicare le soluzioni sviluppate in contesti reali per valutarne l'efficacia e l'efficienza.

### Metodi

Sono state installate delle webcam in Piazza del Popolo, a Pesaro, per monitorare i tempi di permanenza utilizzando **reti neurali YOLOv8**. La libreria Supervision migliora le applicazioni di computer vision e deep learning con funzionalità chiave come il **rilevamento** e **l'annotazione degli oggetti** nei video, il **salvataggio degli oggetti rilevati**, il **filtraggio delle rilevazioni** per migliorare l'accuratezza e l'identificazione efficace di oggetti di piccole dimensioni.

# Use Case: Dwell Time Analysis for People Flow

## Dispositivi Edge

### Efficienza Energetica:

Ideali per monitorare spazi pubblici, funzionando su batterie o risorse energetiche limitate.

### Prestazioni in Tempo Reale:

Cruciali per analizzare i flussi di persone e ridurre i tempi di attesa in tempo reale.

### Scalabilità:

Facilmente distribuibili in diverse aree della città, migliorando la copertura e il monitoraggio.

### Costi Operativi:

Economici in termini di hardware e manutenzione.

## Dispositivi Standard

### Consumo Energetico:

Maggiore, quindi meno adatti per applicazioni distribuite su vasta scala in aree pubbliche.

### Prestazioni:

Offrono maggiore accuratezza, ma con tempi di risposta più lenti se non ottimizzati per l'inferenza in tempo reale.

### Potenza di Calcolo:

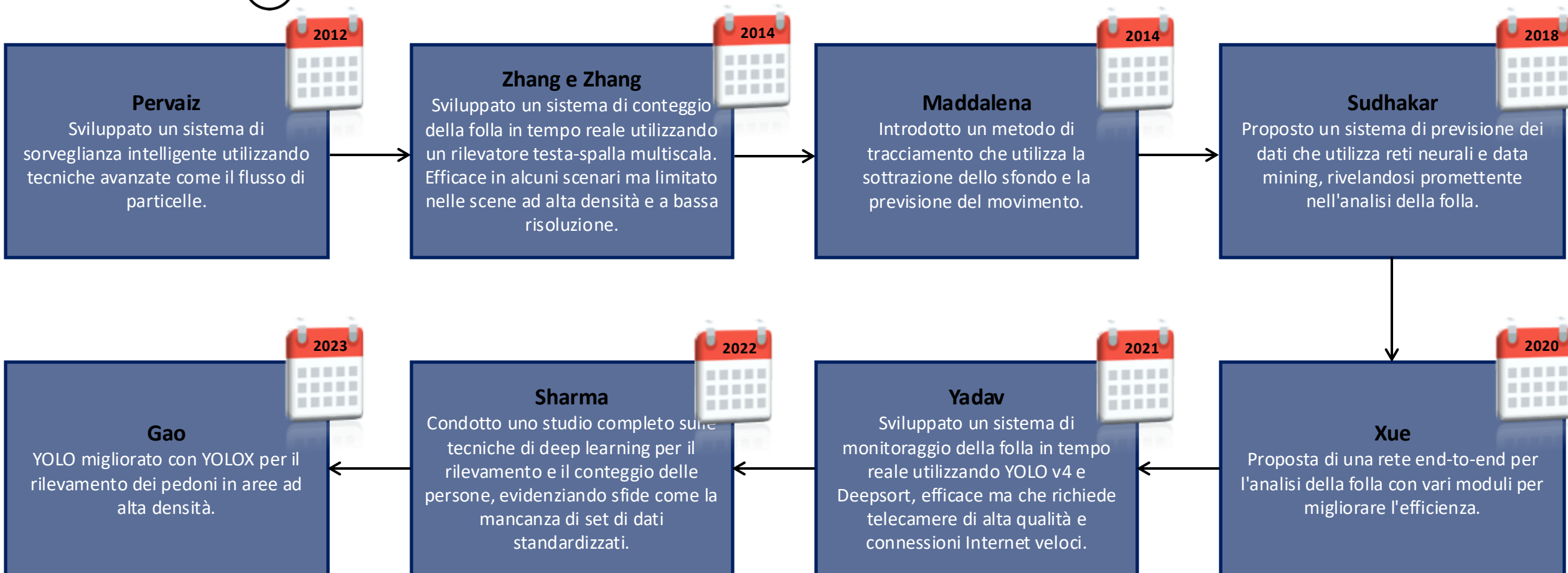
Adatti per addestrare modelli di deep learning complessi per l'analisi avanzata dei flussi di persone.

### Costi:

Elevati per hardware e manutenzione.

# Use Case: Dwell Time Analysis for People Flow

## Stato dell'arte



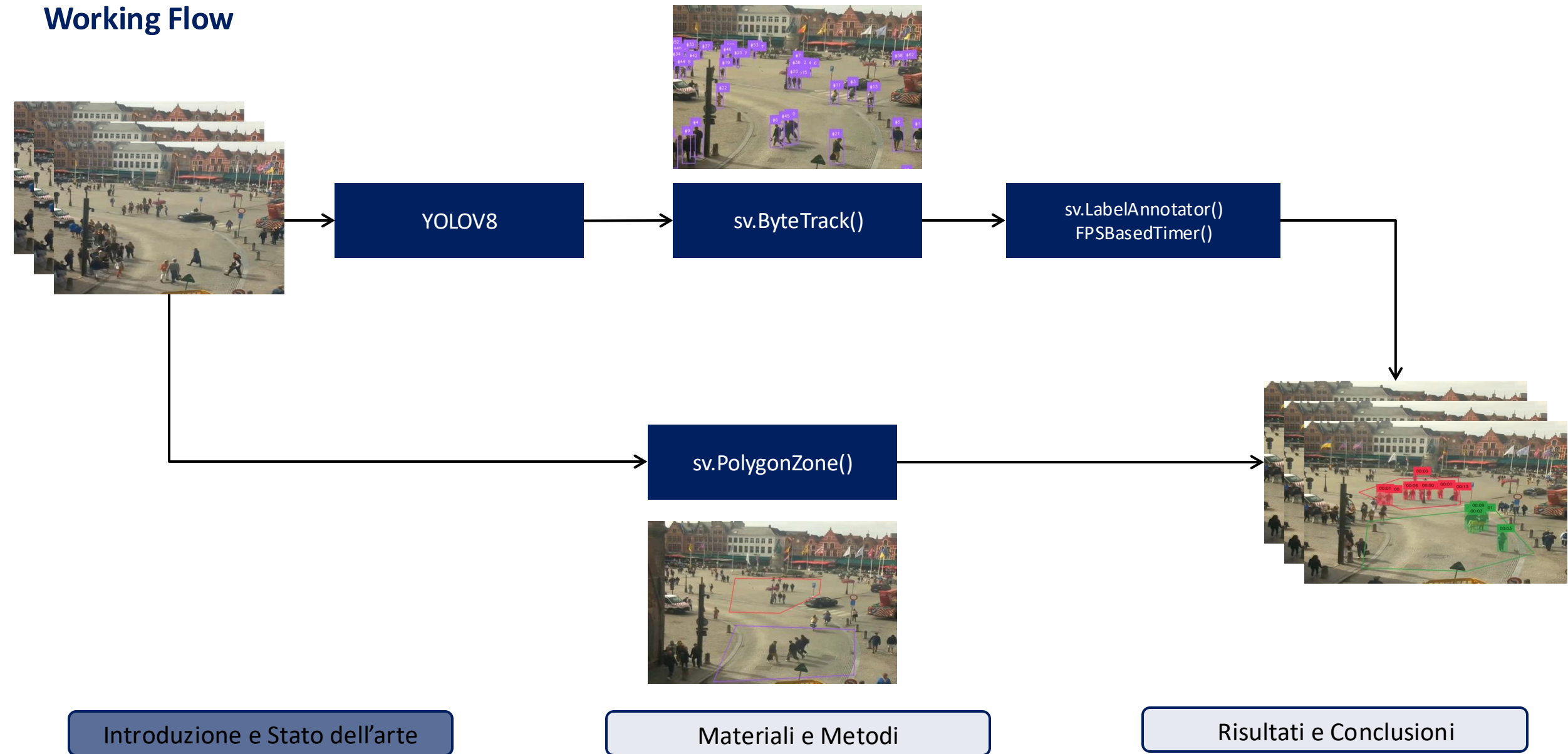
Introduzione e Stato dell'arte

Materiali e Metodi

Risultati e Conclusioni

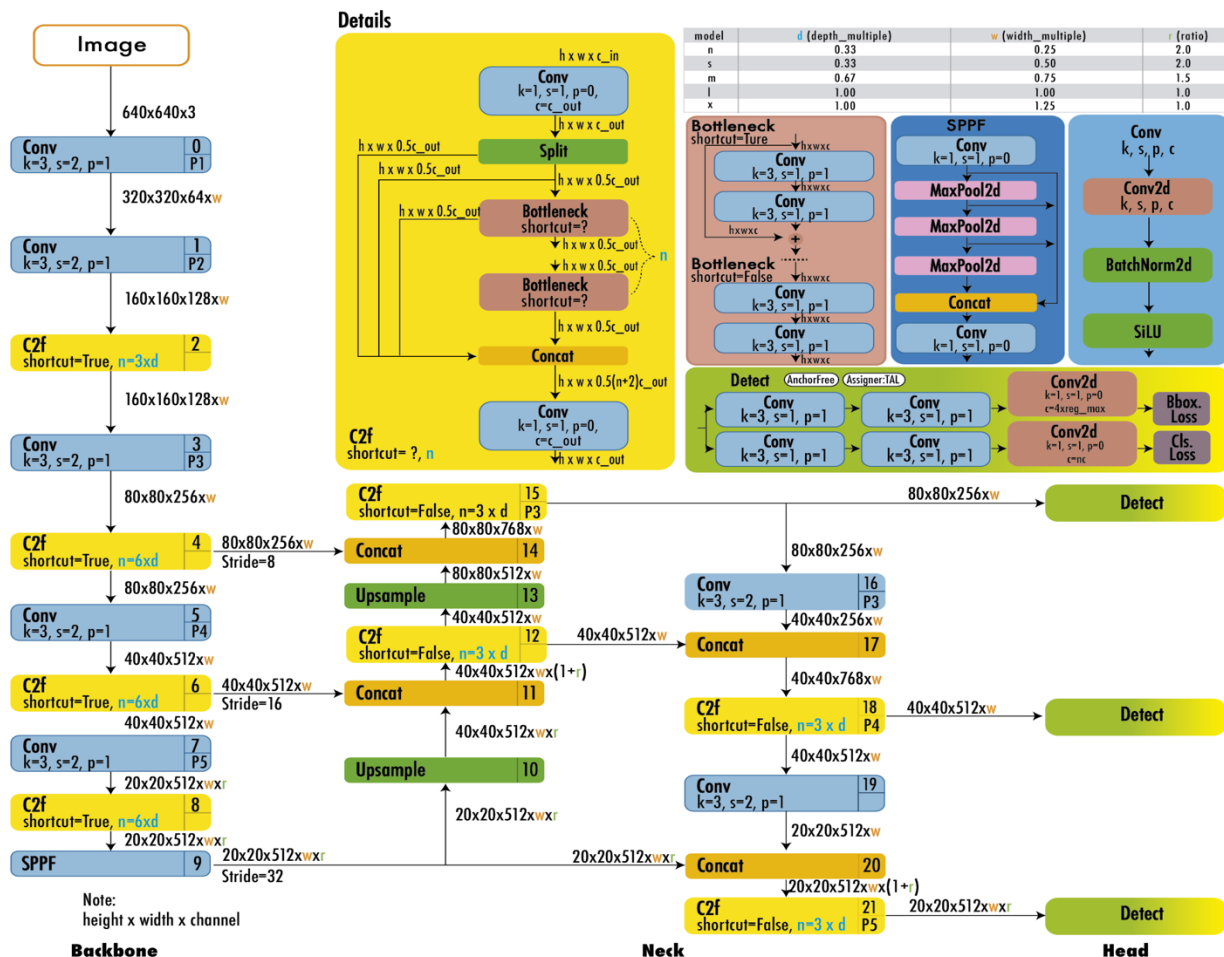
# Use Case: Dwell Time Analysis for People Flow

## Working Flow



# Use Case: Dwell Time Analysis for People Flow

## Architettura YOLOv8



## Innovazioni YOLOv8



### C2f Module

- Sostituisce il modulo C3 delle versioni precedenti, integrando i principi di Cross Stage Partial (CSP) ed Efficient Layer Aggregation Networks (ELAN) per un flusso gradiente più efficiente.

### SPPF (Spatial Pyramid Pooling Fast)

- Mantiene la precisione del target su scale diverse, migliorando la capacità di rilevare oggetti di varie dimensioni.

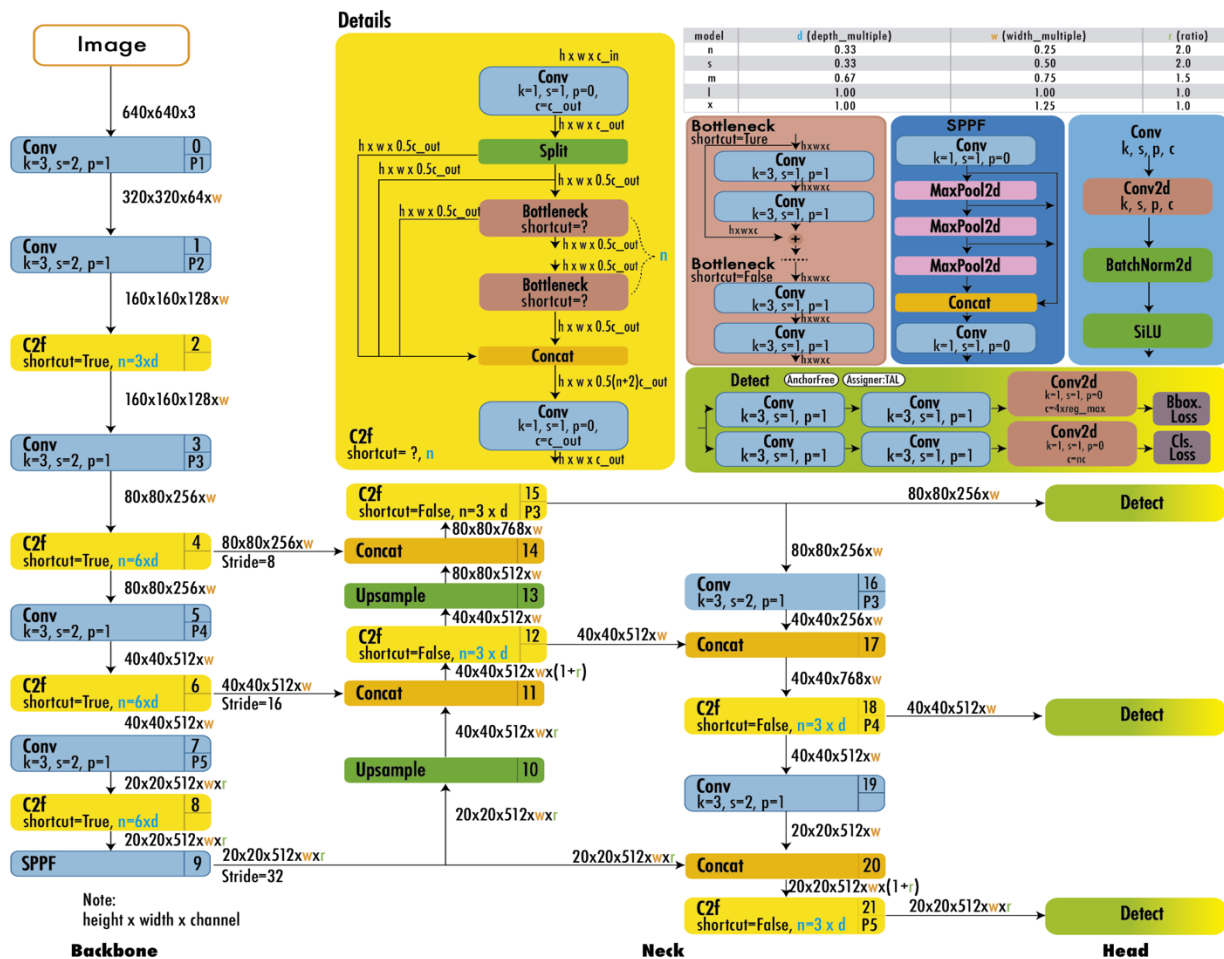
### MaxPool

- Riduce la dimensione delle feature mantenendo le informazioni essenziali, migliorando la precisione del modello.



# Use Case: Dwell Time Analysis for People Flow

## Architettura YOLOv8



## Innovazioni YOLOv8

### Neck PAN-FPN (Path Aggregation Network - Feature Pyramid Network)

- Migliora la fusione di informazioni provenienti da scale diverse, migliorando il rilevamento di oggetti di varie dimensioni.

### Anchor-Free Detection

- YOLOv8 adotta un approccio senza "anchor" per migliorare l'accuratezza delle previsioni degli oggetti, eliminando la necessità di definire "anchor" predefinite.

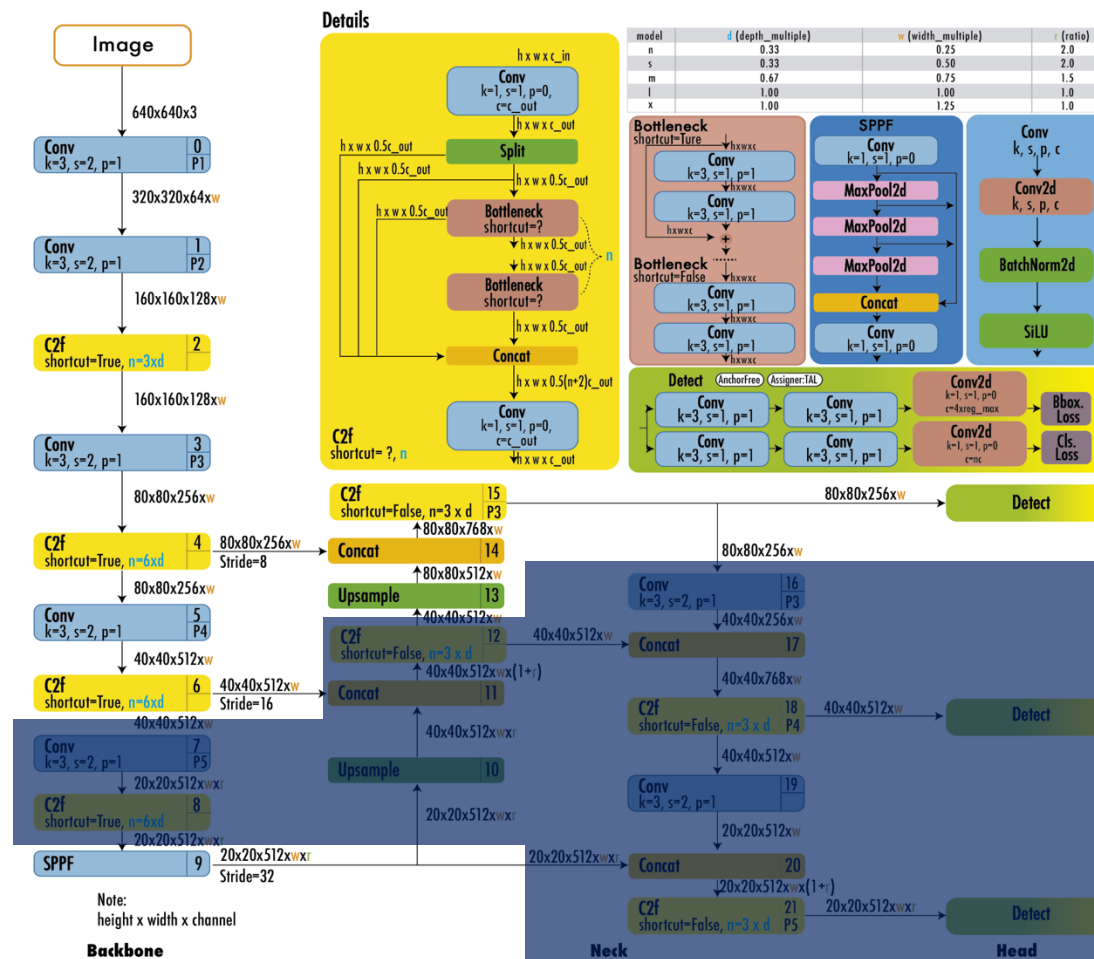
### Optimization Techniques

- Sono state implementate varie modifiche architetturali e algoritmiche per ottimizzare le prestazioni, inclusa una migliore gestione del gradiente e una riduzione del numero di parametri senza compromettere la qualità del rilevamento.



# Use Case: Dwell Time Analysis for People Flow

## Architettura YOLOv8



## Modifiche YOLOv8

Ottimizzazione di YOLOv8 per il rilevamento di persone rimuovendo il rilevamento di oggetti grandi/medi

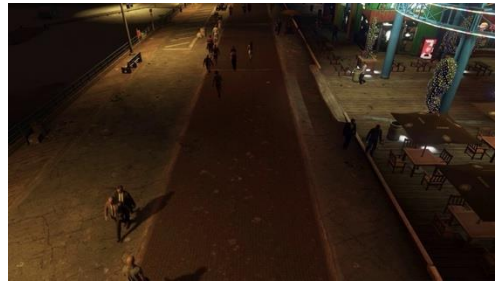
- Per ottimizzare il modello YOLOv8 per il nostro task specifico di rilevamento di persone, considerate di piccola taglia, abbiamo deciso di eliminare la parte relativa al rilevamento di oggetti grandi e medi e mantenere solo il rilevamento di oggetti piccoli.

## Miglioramento delle prestazioni

- Il numero di parametri è stato ridotto da 11,2 milioni a 2,2 milioni, portando ad una significativa diminuzione del carico computazionale della rete, rendendola ideale per applicazioni in tempo reale.
- La velocità di inferenza è stata migliorata rispetto all'architettura classica.

# Use Case: Dwell Time Analysis for People Flow

## Dataset *MOTH-Synth* – Training/Validation set



Ampio dataset **sintetico** creato per il rilevamento dei pedoni.

Sviluppato utilizzando il videogioco altamente fotorealistico **Grand Theft Auto V**, sviluppato da Rockstar North.

### Video

- 764 full-HD videos
- Ogni video contiene 1800 frame
- Registrati a 20 fps

### Totale

- Oltre 1.3 milioni di frame
- 33 milioni di istanze di persone

### Annotazioni Disponibili

- Bounding Boxes
- Maschere di segmentazione delle istanze
- 2D and 3D Keypoints

### Selezionate

- 50 immagini per ogni video dall'alto
- 30,000 istanze di persone
- 8000 immagini dall'alto



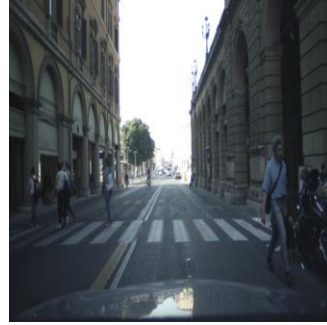
Introduzione e Stato dell'arte

Materiali e Metodi

Risultati e Conclusioni

# Use Case: Dwell Time Analysis for People Flow

## Dataset *EuroCity Persons (ECP)* – Training/Validation set



Si concentra sui pedoni nelle scene del traffico urbano, progettato per il rilevamento dei pedoni.

### Caratteristiche

- 4 stagioni
- Scenari diurni e notturni
- 12 nazioni
- 31 città

### Immagini

- 47,300 immagini
- 238,200 persone annotate

### Annotazioni Disponibili

- Bounding Boxes

### Selezionate

- 6 città italiane
- 3,000 istanze di persone
- 5485 immagini di scene di traffico urbano



# Use Case: Dwell Time Analysis for People Flow

## Dataset *SOMPT22 – Test set*



Si concentra sui pedoni nelle piazze e negli incroci stradali, progettato per il rilevamento dei pedoni

### Annotazioni Disponibili

- Bounding Boxes

### Totale

- 21,602 frame
- 801,341 persone annotate

### Caratteristiche

- 24/7 webcam posizionate a 6 metri di altezza
- 6 città

### Video

- 14 video
- Registrati a 30 fps
- 1 minuto per video

### Selezionate

- 30 frame per video per EuroCity test
- 75 frame per video per MOTH-Synth test



Introduzione e Stato dell'arte

Materiali e Metodi

Risultati e Conclusioni

# Use Case: Dwell Time Analysis for People Flow

## HyperParameters for training

Batch size

- 16

Image size

- 640×640

Number of epochs

- 250

Early stopping patience

- 30

Optimizer

- SGD

Initial learning rate

- 0.01

Final Learning rate

- 0.001

Momentum

- 0.95

IOU threshold

- 0.7

Max detection limit

- 300

Mixed precision (AMP):

- yes

Warmup epochs

- 10

Warmup momentum

- 0.5

Warmup bias learning rate

- 0.1

Weight decay

- 0.0001

## DataAugmentation

Translation augmentation

- 0.5

Scaling

- 0.5

Horizontal flipping

- 0.1

Auto augmentation

- randaugment

Erasing

- 0.4

Mosaic

- 1 in last 10 epochs



# Use Case: Dwell Time Analysis for People Flow

## Risultati dell'addestramento

Dataset	Model	Inference	mAP@50	mAP@50-95	Precision	Recall(R)
MOTSynth	YOLOv8n	1.4ms	0.729	0.430	0.816	0.585
	YOLOv8s	2.1ms	0.749	0.449	0.836	0.606
	YOLOv8m	4.5ms	0.777	0.483	0.849	0.648
	YOLOv8l	7.6ms	0.764	0.470	0.860	0.617
EuroCity Persons	YOLOv8n	1.7ms	0.602	0.401	0.721	0.287
	YOLOv8s	3.7ms	0.629	0.439	0.846	0.298
	YOLOv8m	7.9ms	0.633	0.415	0.844	0.388
	YOLOv8l	13.5ms	0.621	0.425	0.900	0.328
MOTSynth & Eurocity	YOLOv8n	1.2ms	0.741	0.454	0.824	0.6
	YOLOv8s	2.2ms	0.778	0.485	0.850	0.651
	YOLOv8m	4.6ms	0.781	0.488	0.854	0.652
	YOLOv8l	7.6ms	0.786	0.487	0.859	0.659
Coco	YOLOv8n	1.2ms	0.656	0.434	0.909	0.377
	YOLOv8s	2.2ms	0.691	0.459	0.920	0.447
	YOLOv8m	4.6ms	0.699	0.471	0.919	0.463
	YOLOv8l	7.9ms	0.683	0.465	0.919	0.431

L'addestramento delle varianti YOLOv8 su set di dati specifici come MOTHSynth, EuroCity Persons e il set di dati combinato MOTHSynth ed EuroCity ha portato a notevoli miglioramenti delle prestazioni rispetto ai modelli pre-addestrati su COCO.

### • Miglioramenti in Recall

- La maggior parte delle varianti addestrate su set di dati specifici hanno mostrato un aumento in precision e recall. Ad esempio, YOLOv8m addestrato su MOTHSynth ed EuroCity ha registrato una precision di 0,854 e un recall di 0,652, rispetto a 0,919 e 0,463 per il modello pre-addestrato su COCO, indicando una riduzione dei falsi positivi e un aumento dei veri positivi.

### • Miglioramenti in mAP

- Le varianti YOLOv8 addestrate su set di dati specifici hanno generalmente ottenuto valori mAP@50 e mAP@50-95 più elevati. Ad esempio, YOLOv8l su MOTHSynth ed EuroCity ha raggiunto un mAP@50 di 0,786 e un mAP@50-95 di 0,487, rispetto a 0,683 e 0,465 per il modello COCO, dimostrando una maggiore precisione complessiva.

### • Velocità di Inferenza

- Sebbene leggermente più lente, le varianti addestrate su set di dati specifici mantengono una velocità di inferenza competitiva. La velocità degli YOLOv8 su MOTHSynth ed EuroCity è di 2,2 ms, un compromesso accettabile visti i miglioramenti delle prestazioni.

# Use Case: Dwell Time Analysis for People Flow

## Il Modello Migliore

Dataset	Model	Inference	mAP@50	mAP@50-95	Precision	Recall(R)
MOTSynth	YOLOv8n	1.4ms	0.729	0.430	0.816	0.585
	YOLOv8s	2.1ms	0.749	0.449	0.836	0.606
	YOLOv8m	4.5ms	0.777	0.483	0.849	0.648
	YOLOv8l	7.6ms	0.764	0.470	0.860	0.617
EuroCity Persons	YOLOv8n	1.7ms	0.602	0.401	0.721	0.287
	YOLOv8s	3.7ms	0.629	0.439	0.846	0.298
	YOLOv8m	7.9ms	0.633	0.415	0.844	0.388
	YOLOv8l	13.5ms	0.621	0.425	0.900	0.328
MOTSynth & Eurocity	YOLOv8n	1.2ms	0.741	0.454	0.824	0.6
	YOLOv8s	2.2ms	0.778	0.485	0.850	0.651
	YOLOv8m	4.6ms	0.781	0.488	0.854	0.652
	YOLOv8l	7.6ms	0.786	0.487	0.859	0.659
Coco	YOLOv8n	1.2ms	0.656	0.434	0.909	0.377
	YOLOv8s	2.2ms	0.691	0.459	0.920	0.447
	YOLOv8m	4.6ms	0.699	0.471	0.919	0.463
	YOLOv8l	7.9ms	0.683	0.465	0.919	0.431

## YOLOv8s: ideale per applicazioni di monitoraggio delle piazze

YOLOv8s su **MOTH-Synth** ed **Eurocity** rappresenta un compromesso ottimale per la nostra applicazione del caso d'uso, offrendo:

- **Velocità di inferenza:** Veloce (2.2ms), cruciale per le applicazioni in tempo reale.
- **Precision:** Alta (0.85), garantendo pochi falsi positivi.
- **Recall:** Alto (0.651), garantendo che la maggior parte dei pedoni venga rilevata correttamente.
- **mAP@50 and mAP@50-95:** Valori rispettivamente di 0,778 e 0,485, indicano un'elevata precisione complessiva.

L' addestramento su un dataset combinato che include **persone reali** (EuroCity Persons) e **sintetiche** (MOTH-Synth) può portare a un modello più robusto e preciso.



# Use Case: Dwell Time Analysis for People Flow

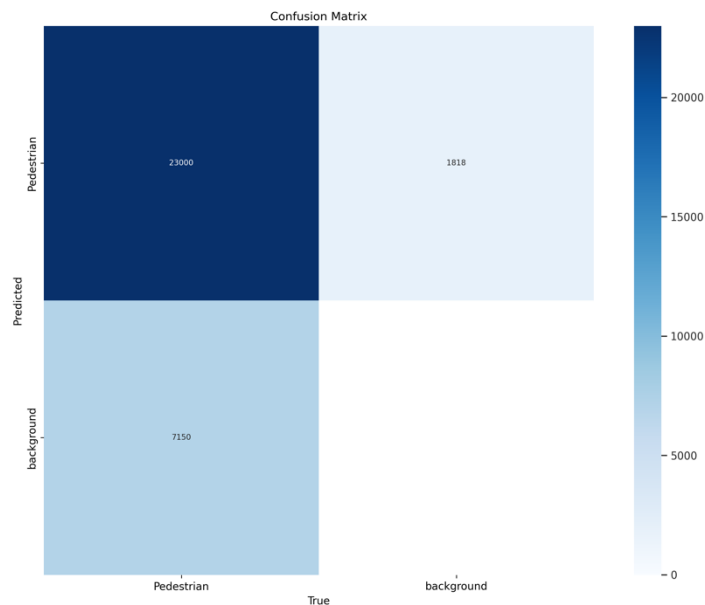
## Tra dati sintetici e reali

Utilizzare sia **dataset reali** che **sintetici** è stato **cruciale** per raggiungere risultati eccellenti.

- A causa delle limitazioni imposte dalla **privacy**, la disponibilità di dati adatti al monitoraggio dall'alto è stata ridotta. In questo contesto, MOTH-Synth è stato particolarmente prezioso, permettendo la simulazione di scenari urbani senza compromettere la privacy.
- YOLOv8 si è rivelata capace di rilevare pedoni utilizzando immagini sintetiche, dimostrando **robustezza** e **flessibilità**, specialmente in situazioni in cui i dati reali erano scarsi.
- Questo approccio **ibrido**, che combina l'accuratezza dei dati reali con la versatilità dei dati sintetici, ha creato un **equilibrio ideale**.
- I **migliori risultati** sono stati ottenuti addestrando il modello su una combinazione di dataset reali e sintetici, portando a un modello **robusto** e **capace di generalizzare efficacemente** in vari scenari.

# Use Case: Dwell Time Analysis for People Flow

## YOLOv8s - "Analisi grafica dei risultati"



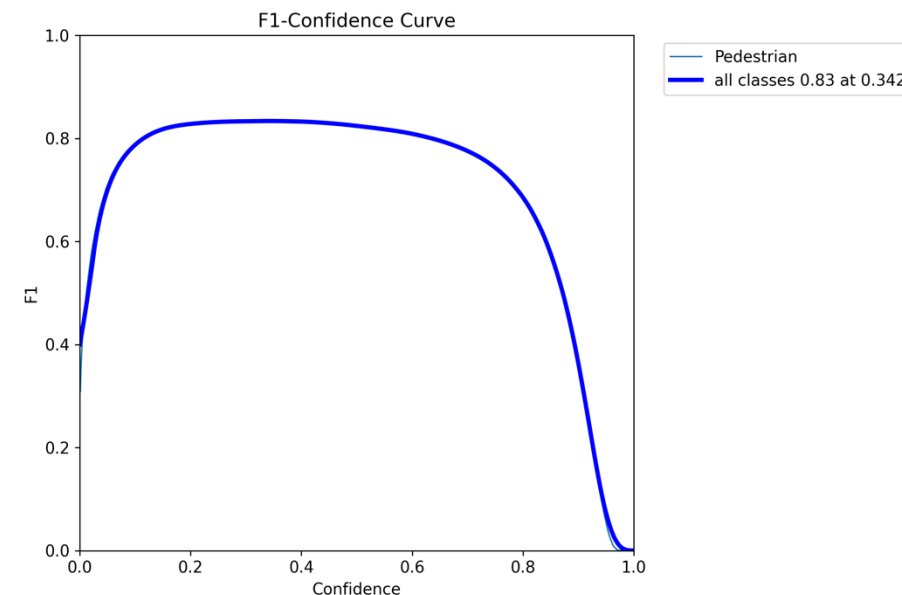
### Matrice di Confusione

La matrice di confusione mostra buona precision nel rilevamento dei pedoni, con **23.000 true positives e 7.150 false negatives**. Tuttavia, ci sono **1.818 false positives**, indicando che il modello tende a identificare gli elementi di background come pedoni.

### F1-Confidence Curve

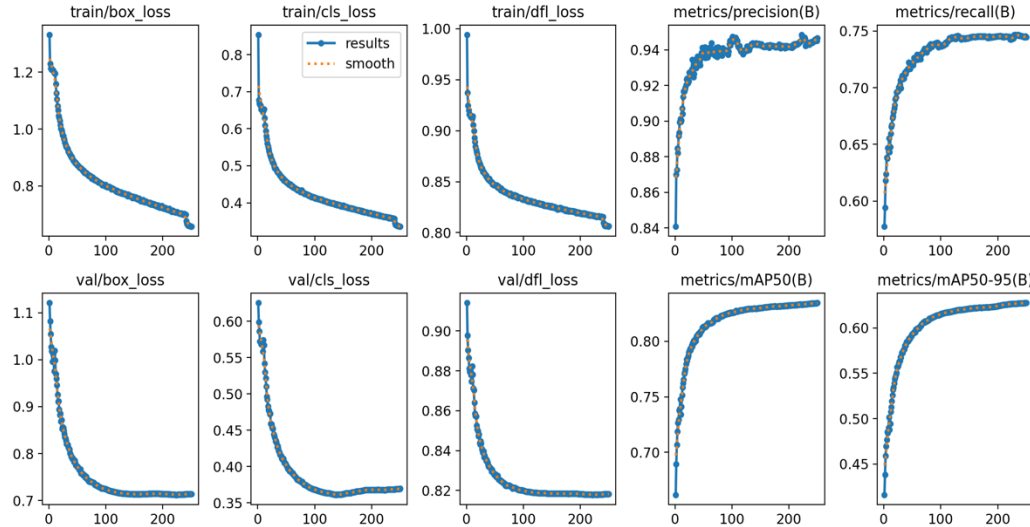
La curva F1, tracciata rispetto ai livelli di confidenza, mostra che il punteggio **F1** ottimale è **0,82** a un **livello di confidenza** di **0,393**.

Ciò suggerisce che il modello mantiene un buon equilibrio tra **precision** e **recall** a questo **livello di confidenza**.



# Use Case: Dwell Time Analysis for People Flow

## YOLOv8s - "Analisi grafica dei risultati"

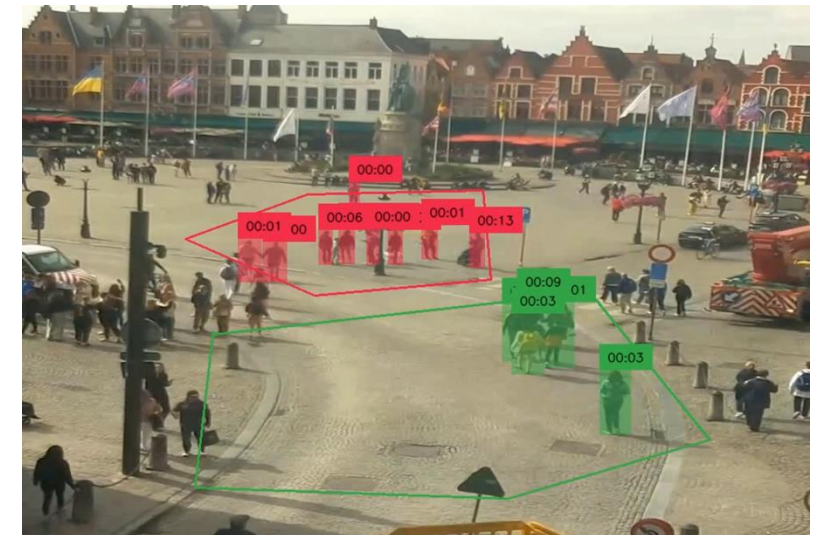


### Metriche di Addestramento

I grafici delle metriche di addestramento mostrano un miglioramento costante della precision e del recall nel tempo, con una precision che si stabilizza intorno a 0,95 e un recall a 0,74. Le loss di addestramento e validazione diminuiscono costantemente, indicando un processo di addestramento ben funzionante.

### Risultato finale

I risultati ottenuti sono promettenti nel campo del monitoraggio dei flussi di persone attraverso l'utilizzo di tecniche avanzate di Computer Vision e Deep Learning, dimostrando che l'approccio utilizzato nel progetto è valido e offre grandi potenzialità per l'ottimizzazione degli spazi pubblici.



## Risultati di YOLOv8s modificata per il rilevamento di piccoli oggetti

Questa configurazione è stata valutata utilizzando un set di test composto da 1005 immagini dal dataset SOMPT22.

- **Box Precision (P)**
  - 0.826
- **Recall (R)**
  - 0.641
- **mAP@50**
  - 0.766
- **mAP@50-95**
  - 0.47

Significativo miglioramento della velocità nell'inferenza:

- 
- Figure 10 displays two side-by-side images of a street scene with pedestrians, comparing detection results. The left image is labeled 'FPS: 29.68' and the right image is labeled 'FPS: 46.91'. Both images show multiple pedestrians with bounding boxes and confidence scores. The left image has a higher FPS but the right image has a higher FPS and more detections.

- **Classica**
  - 11.2 milioni
- **Modificata**
  - 2.2 milioni

# Use Case: Dwell Time Analysis for People Flow

## Conclusioni

Il progetto "Dwell Time Analysis for People Flow" ha dimostrato l'efficacia di tecniche avanzate di Computer Vision e Deep Learning per il monitoraggio del flusso di persone e l'ottimizzazione degli spazi pubblici, con YOLOv8 che si è rivelata particolarmente efficace per il monitoraggio in tempo reale.

## Sviluppi Futuri

- **Testing su Video Streaming**
  - Ottenere video in tempo reale per convalidare il sistema.
- **Training con Dataset più grandi**
  - Utilizzare dataset più grandi per migliorare la generalizzazione del modello.
- **Ottimizzazione per dispositivi Edge**
  - Migliorare l'efficienza computazionale di YOLOv8 per applicazioni in tempo reale.
- **Integrazione di dataset**
  - Addestrare il modello con dataset che rappresentano scenari urbani diversi e complessi.

# Use Case: Dwell Time Analysis for People Flow

## Impatto ambientale dell'addestramento di YOLOv8s

L'addestramento di modelli di deep learning come YOLOv8 richiede una notevole quantità di risorse computazionali, contribuendo in modo significativo all'impronta di CO2.



## Emissioni di CO2 e consumo di elettricità dell'addestramento di YOLOv8s

Consumo Elettrico : **3.755 kW/h**



CO2 Emissioni di CO2 : **1.42 kg CO2eq**



Ciò è equivalente a:

5,6 km percorsi  
con un'auto a  
benzina



0,74 kg di  
carbone bruciato



0,067 alberi  
piantati e  
cresciuti da 10  
anni



# Use Case: Dwell Time Analysis for People Flow

## Link Utili

- Modelli Addestrati



- Dataset



- Repository GitHub



## Ringraziamenti

Un ringraziamento va ai nostri **tutor** per averci aiutato durante tutto il progetto.

Infine, volevamo ringraziarvi per la vostra attenzione e il vostro tempo.

Andrea Faccenda – S1119359  
Simone Di Battista - S1120038



UNIVERSITÀ  
POLITECNICA  
DELLE MARCHE

