UNIVERSITÀ POLITECNICA DELLE MARCHE

FACOLTÀ DI INGEGNERIA

Corso di Laurea triennale in **Ingegneria Informatica e dell'Automazione**

Elaborazione e classificazione di flussi RGB-D

Processing and classification of RGB-D streams

Relatore:
Prof. ***Gambi Ennio***

Correlatore:
Prof.ssa ***Senigagliesi Linda***

Tesi di Laurea di:
Ciuffreda Luigi

*Matricola:****1100762***

Anno Accademico 2022/2023

A chi ha creduto in me fin dall'inizio.
E fino alla fine.

Abstract

Questo lavoro di tesi si focalizza sull'elaborazione e la classificazione dei flussi RGB-D acquisiti tramite la fotocamera Intel RealSense D455. L'obiettivo principale è valutare l'adattabilità di algoritmi precedentemente addestrati su immagini RGB al contesto dei dati di profondità. Attraverso sessioni di acquisizione dati, dove i partecipanti assumono una varietà di pose in diversi ambienti, abbiamo raccolto dati rappresentativi della vita reale. Durante lo studio, sono stati condotti tre cicli di addestramento e test su dati di profondità, seguendo un approccio di allenamento incrociato. Nei primi due cicli, il modello è stato addestrato su dati provenienti da due background diversi e testato su un terzo background. Nel terzo ciclo, il modello è stato nuovamente addestrato su dati provenienti dai primi due background e testato su un terzo background differente. L'accuratezza media del modello durante l'addestramento ha raggiunto il 95% sul dataset di addestramento. Durante il testing su dati di profondità provenienti da ambienti con sfondi diversificati, il modello ha mostrato un'accuratezza media del 90% per tutti e tre i cicli. La capacità del modello di adattarsi a contesti ambientali variabili è stata evidente grazie ai valori di recall medi superiori al 95% in tutti e tre i test. In conclusione, questo studio ha dimostrato che gli algoritmi addestrati su immagini RGB possono essere applicati con successo a dati di profondità in contesti reali. Tuttavia, è importante notare che la precisione rimane bassa, con valori medi del 25%. La recall, d'altra parte, è pari a 1, indicando la capacità del modello di riconoscere tutte le istanze di persone. La F1-score si attesta intorno al 40-50%, suggerendo che il modello sta ottenendo un equilibrio tra recall e precision, ma c'è ancora spazio per miglioramenti.

Indice

Elenco delle figure

Capitolo 1

Introduzione

Nel presente lavoro di tesi, si introduce un approccio sistematico per il riconoscimento delle persone attraverso dati di profondità acquisiti mediante l'utilizzo della fotocamera RGB-D Intel RealSense D455. L'obiettivo primario di questa ricerca è esplorare e analizzare la capacità e l'efficacia di algoritmi precedentemente addestrati su immagini RGB,come descritto in questo articolo [1], in particolare YOLO, nell'interpretare e processare immagini di profondità. Inoltre, è importante notare che gli algoritmi che operano direttamente su immagini di profondità presentano meno preoccupazioni in merito alla privacy e possono funzionare efficacemente anche in condizioni di scarsa illuminazione, come di notte. Inoltre, l'immagine di profondità è bidimensionale, mentre gli approcci basati su immagini RGB operano in tre dimensioni. Pertanto, lo sviluppo di algoritmi che lavorano con immagini di profondità offre il vantaggio di una significativa riduzione dei costi computazionali. In dettaglio, diversi partecipanti sono stati coinvolti in sessioni di acquisizione durante le quali hanno assunto una varietà di posture, consentendo di registrare una gamma complessa e variegata di dati. Questi dati, acquisiti in ambienti con differenti sfondi, sono stati poi utilizzati come input per i suddetti algoritmi, con l'intento di valutare la loro accuratezza e affidabilità nel riconoscimento delle persone attraverso immagini di profondità. È fondamentale sottolineare che gli algoritmi selezionati, originariamente addestrati e ottimizzati per lavorare con immagini RGB, sono stati applicati in questo studio con processi di addestramento su pochi dati di profondità. La ricerca mira, quindi, a indagare la generalizzabilità e la trasferibilità delle competenze apprese da questi algoritmi su dati RGB, quando applicati a dati di profondità, offrendo così un'analisi approfondita e un'accurata valutazione delle loro prestazioni in tale contesto applicativo inedito.

Capitolo 2

Stato dell'arte

Nel contesto della mia ricerca, ho impiegato una serie di algoritmi avanzati per l'elaborazione dell'immagine RGB. Questi algoritmi, noti come MediaPipe, YOLO Face e Cascade Detection, hanno costituito una parte fondamentale della metodologia utilizzata per l'analisi delle immagini acquisite.

2.1 Mediapipe [2]

MediaPipe è stata la prima libreria impiegata nel mio progetto. Sviluppata da Google, MediaPipe offre un ampio spettro di funzionalità di elaborazione dell'immagine, tra cui il rilevamento delle pose umane, delle mani e dei volti. L'integrazione di MediaPipe si è dimostrata cruciale per identificare e monitorare le posizioni delle persone nell'immagine catturata dalla telecamera. Questa funzionalità ha trovato applicazione in diversi contesti, come l'analisi del movimento umano.

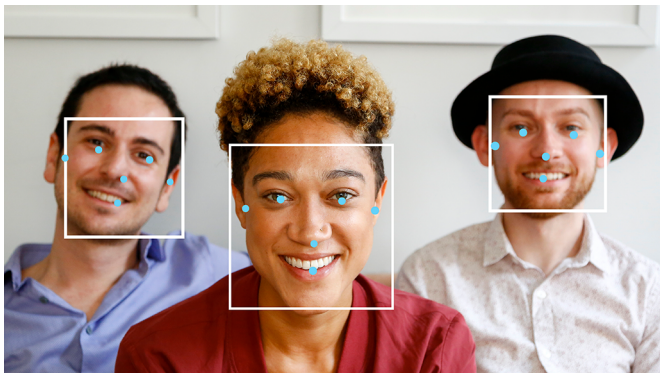


Figura 2.1: Mediapipe

2.2 YOLO Face [3]

YOLO Face, invece, è noto per la sua abilità nel rilevare volti umani in immagini e video. Nel corso della mia ricerca, ho incorporato con successo l'algoritmo YOLO Face per individuare e analizzare i volti presenti nell'immagine proveniente dalla telecamera. Tale funzionalità si è dimostrata di grande rilevanza nelle applicazioni legate all'identificazione e al monitoraggio delle persone.

Figura 2.2: Yolo Face

2.3 Cascade Classifier [4]

Inoltre, ho utilizzato l'algoritmo di rilevamento a cascata (Cascade Classifier) per affrontare specifici compiti di rilevamento di persone e caratteristiche nelle immagini acquisite. L'integrazione di questo algoritmo ha permesso di ottenere risultati accurati e affidabili nell'identificazione di determinati pattern e strutture all'interno delle immagini.

Figura 2.3: Cascade Classifier

L'impiego combinato di questi algoritmi ha consentito di affrontare in modo esaustivo le sfide legate all'elaborazione dell'immagine dalla telecamera RGB nel contesto della mia ricerca. La loro integrazione ha fornito una solida base per l'analisi e l'interpretazione delle informazioni visive raccolte durante il corso dello studio.

2.4 Alti approcci simili in letteratura scientifica

Nella tabella sottostante 2.1, sono stati inseriti alcuni articoli simili trovati nella letteratura, ognuno con il suo contributo principale, vantaggi, svantaggi, dati di input, tipologia di algoritmo utilizzato e dataset associato:

Questi articoli forniscono un'ampia panoramica delle diverse ricerche condotte nell'ambito del rilevamento di oggetti, persone e ostacoli utilizzando varie metodologie e approcci. Le informazioni presentate nella tabella consentono di confrontare le diverse caratteristiche di ciascun articolo e di valutarne l'applicabilità e le prestazioni in base alle specifiche esigenze di ricerca o di sviluppo.Questo elenco offre una visione completa degli sforzi precedenti nel campo, fornendo una base solida per la discussione e il confronto con i risultati e le scoperte presentati in questo studio.

NOME ARTICOLO	CONTRIBUTO PRINCIPALE	VANTAGGI	SVANTAGGI	SCOPO	INPUT	ALGORITMO	PERFORMANCE	DATASET
Expandable YOLO: 3D Object Detection from RGB-D Images [5]	Rilevazione efficace	Leggero	Limitazioni del Dataset	Object Detection	RGB-D	YOLO	Intersection over Union	PRIVATO
LeggerTowards Real-time 3D Object Detection for Autonomous Mobile Robots in Logistics Scenarios [6]	Focus su Scenari Logistici	Applicazioni Pratiche	Mancanza di Risultati Dettagliati	Object Detection	RGB-D	DL	Recall 94%	KITTI[7]
Cross-Modal Analysis of Human Detection for Robotics: An Industrial Case Study [8]	Strategia di Transfer Learning	Risultati Variabili	Necessità di Ulteriori Ricerche	Human Detection	RGB-D	DL	High Precision	KITTI[7]
Accurate detection and 3-D localization of humans using a new RGB-D fusion approach based on YOLO and synthetic training data [9]	Testato su diversi dataset	Precisione del 96,5%	Richiede numerosi dati per l'addestramento	Human Detection	RGB-D	YOLO v3	High Precision	Privato
An Algorithm for Obstacle Detection based on YOLO and Light Filed Camera [10]	Focus su immagini RGB.D	Elevata Accuratezza	Limitazioni del Dataset	Obstacle Detection	RGB-D	YOLO	High Accuracy	Privato
Fast heuristic method to detect people in frontal depth images [11]	Focus su immagini RGB-D	Elevata Precision	Necessità ulteriori ricerche	Human Detection	RGB-D	DL	Precision of 99.26%	GFPD[12]
Depth-Based Human Detection Considering Postural Diversity and Depth Missing in Office Environment[13]	Adattamento delle caratteristiche	Elevata Accuratezza	Limitazioni hardware	Human Detection	DEPTH	ML	97.7 % Accuracy	Privato
IPT: A Dataset for Identity Preserved Tracking in Closed Domains[14]	Valutazione del Tracking	Dataset Diversificato	Complessità del Dataset	Identity Preserved Tracking	DEPTH	YOLO v3	High Performance	Aircraft Context [15].
Depth edge detection using edge-preserving filter and morphological operations [16]	Nuovo Metodo di Edge Detection	Metodo Innovativo	Specificità del Contesto	Edge detection	DEPTH	ML	High Performance	Middlebury[17] , NYUDepthV2[18]
Towards Silhouette-Aware Human Detection in Depth Images[19]	Utilizzo delle Silhouettes	Privacy	Limitazioni del Dataset	Human Detection	DEPTH	DL	High Accuracy	Pubblico [20]
TIMo—A Dataset for Indoor Building Monitoring with a Time-of-Flight Camera [21]	Valutazione Sperimentale	Privacy	Specificità delle Applicazioni	Human Detection	DEPTH	DL	High Precision	TIMo[22]

Tabella 2.1: Tabella Articoli Simili

Nel primo articolo intitolato 'Expandable YOLO: 3D Object Detection from RGB-D Images [5]'. Il contributo principale di questo lavoro è la realizzazione di un sistema di rilevazione di oggetti efficace che sfrutta sia informazioni sul colore (RGB) che sulla profondità (D) per identificare e localizzare gli oggetti nell'ambiente. Tra i vantaggi di questo approccio si annovera la leggerezza computazionale del modello, che lo rende adatto anche per applicazioni in tempo reale. Tuttavia, è importante notare che questo metodo presenta alcune limitazioni legate al dataset utilizzato per l'addestramento, che potrebbero influire sulla sua capacità di generalizzazione. Lo scopo principale di questo lavoro è l'implementazione di un sistema di rilevazione degli oggetti, con un focus specifico sull'ambito della "Object Detection". Gli input utilizzati sono dati RGB-D, e l'algoritmo principale impiegato è YOLO (You Only Look Once). Le prestazioni del modello sono valutate utilizzando l'indice di "Intersection over Union", e va notato che il dataset utilizzato per la valutazione è privato, il che potrebbe limitare la possibilità di confronto con altri approcci presenti in letteratura.

La riga numero due "Towards Real-time 3D Object Detection for Autonomous Mobile Robots in Logistics Scenarios" [6] rappresenta un articolo di ricerca focalizzato sull'implementazione di un sistema di rilevazione tridimensionale di oggetti in tempo reale per l'utilizzo su robot mobili autonomi in scenari logistici. Il contributo principale di questo lavoro è la progettazione di un sistema di rilevazione degli oggetti specificamente mirato a contesti logistici, il che ne aumenta l'applicabilità pratica. Questo approccio promette di essere utile per applicazioni pratiche nel settore della robotica, tuttavia, una limitazione importante è la mancanza di dettagli completi sui risultati ottenuti. L'obiettivo principale di questa ricerca è la rilevazione degli oggetti, con un'enfasi specifica sull'ambito della "Object Detection". Gli input utilizzati consistono in dati RGB, e l'algoritmo principale implementato è il Deep Learning (DL). Le prestazioni del sistema

sono valutate utilizzando il parametro di ”Recall” con un valore del 94%, indicando la capacità del sistema di identificare correttamente la maggior parte degli oggetti rilevanti. È importante notare che il dataset utilizzato per la valutazione è il KITTI [7], che rappresenta uno standard nel campo della rilevazione degli oggetti in scenari di guida autonoma.

La riga successiva riguarda uno studio chiamato ”Cross-Modal Analysis of Human Detection for Robotics: An Industrial Case Study” [8]. Il contributo principale di questo studio è una strategia di apprendimento trasferito (Transfer Learning) che viene utilizzata per rilevare la presenza di esseri umani in ambienti industriali. Tra i vantaggi di questo approccio ci sono risultati variabili, il che suggerisce una certa flessibilità nell'applicazione. Tuttavia, vi è una limitazione chiara, in quanto sono necessarie ulteriori ricerche per affinare e migliorare ulteriormente questo metodo. Lo scopo principale è la rilevazione di persone, mentre l'input utilizzato è basato su dati RGB-D. L'algoritmo impiegato è il Deep Learning (DL), che è noto per la sua efficacia in compiti di rilevamento. Le prestazioni riportate sono caratterizzate da un'alta precisione, il che suggerisce una bassa probabilità di falsi positivi. Il dataset utilizzato per valutare il metodo è il KITTI[7].

La riga numero quattro descrive uno studio intitolato ”Accurate detection and 3-D localization of humans using a new RGB-D fusion approach based on YOLO and synthetic training data”[9]. Il contributo principale è un nuovo metodo che utilizza dati RGB-D e YOLO v3 per rilevare e localizzare con alta precisione persone in 3D. Tuttavia, richiede una grande quantità di dati di addestramento e il dataset utilizzato è privato.

La riga numero cinque presenta uno studio denominato ”An Algorithm for Obstacle Detection based on YOLO and Light Filed Camera” [10]. Il contributo principale di questo lavoro è un algoritmo che si concentra sul rilevamento di ostacoli utilizzando immagini RGB e YOLO. L'algoritmo offre elevata accuratezza, ma presenta limitazioni legate al dataset utilizzato, che è privato.

La riga numero sei denominata 'Fast heuristic method to detect people in frontal depth images'[11], ci evidenzia il contributo principale dell'articolo nell'offrire un metodo euristico veloce per rilevare persone in immagini di profondità frontale. Questo approccio presenta notevoli vantaggi, tra cui un'alta precisione del 99,26%. Tuttavia, è importante notare che potrebbero essere necessarie ulteriori ricerche per ottimizzare ulteriormente il metodo. Il suo scopo principale è il rilevamento umano, utilizzando immagini RGB come input e sfruttando un algoritmo basato su deep learning. L'articolo fa riferimento al dataset GFPD [12] per valutare le performance del metodo.

La riga successiva invece,intitolata 'Depth-Based Human Detection Considering Postural Di-

versity and Depth Missing in Office Environment' [13], si evince il contributo principale è un metodo di rilevamento umano basato sulla profondità, che tiene conto della diversità posturale e delle mancanze di dati di profondità in un ambiente d'ufficio. L'approccio offre un'accuratezza del 97,7% ma presenta limitazioni legate all'hardware utilizzato. Il metodo utilizza caratteristiche adattate e si basa su machine learning con input RGB.

Nell'articolo 'IPT: A Dataset for Identity Preserved Tracking in Closed Domains'[14], il contributo principale riguarda la valutazione del tracking di identità preservata in domini chiusi. Il dataset è diversificato ma presenta una complessità significativa. Il tracking si basa sull'utilizzo di YOLO v3 e offre prestazioni elevate, utilizzando il dataset Aircraft Context[15].

Nell'articolo 'Depth edge detection using edge-preserving filter and morphological operations'[16], il contributo principale è un nuovo metodo innovativo di edge detection basato sull'utilizzo di filtri conservativi dei bordi e operazioni morfologiche. Questo metodo offre elevate performance ed è applicabile a contesti specifici. Utilizza machine learning con input RGB e si basa sui dataset Middlebury[17] e NYUDepthV2[18].

Nell'articolo 'Towards Silhouette-Aware Human Detection in Depth Images'[19], il contributo principale è l'utilizzo delle silhouette per il rilevamento umano in immagini di profondità. L'approccio offre privacy, ma presenta limitazioni legate al dataset utilizzato. Si basa su deep learning con input RGB e offre un'alta accuratezza. Il dataset utilizzato è pubblico.[20]

Infine, nell'ultima riga della tabella, nell'articolo intitolato 'TIMo—A Dataset for Indoor Building Monitoring with a Time-of-Flight Camera'[21], dove il contributo principale riguarda la valutazione sperimentale per il monitoraggio degli edifici interni utilizzando una telecamera time-of-flight. L'approccio offre privacy ma è specifico per determinate applicazioni. Il rilevamento umano si basa su deep learning con input RGB e offre un'alta precisione grazie anche al dataset TIMo [22]

Capitolo 3

Materiali e Metodi

3.1 Intel RGB-D [23]

La Intel RealSense D455 è una fotocamera avanzata di profondità che impiega una tecnologia di visione stereo per fornire dati tridimensionali accurati e ad alta risoluzione. Equipaggiata con due sensori di profondità e un sensore RGB, questo dispositivo è progettato per offrire prestazioni eccellenti in una varietà di applicazioni. I sensori di profondità operano con una risoluzione di 1280x800 pixel, permettendo la cattura di dettagli fini e la generazione di immagini di profondità dettagliate e nitide. Il sensore RGB, che supporta una risoluzione Full HD di 1920x1080 pixel, consente di acquisire immagini a colori ad alta definizione, che possono essere allineate con i dati di profondità per creare immagini RGB-D. Questo aspetto è fondamentale per applicazioni che richiedono una rappresentazione dettagliata dell'ambiente, oltre all'informazione di profondità. In termini di frame rate, la D455 è in grado di registrare dati di profondità a un massimo di 90 frame per secondo (fps), garantendo una cattura di movimento fluida e reattiva, essenziale per applicazioni in tempo reale. Il frame rate elevato facilita la tracciabilità accurata degli oggetti in movimento e migliora la capacità del dispositivo di funzionare in ambienti dinamici e in rapida evoluzione.

Figura 3.1: Intel D455

La tecnologia di funzionamento si basa sulla visione stereoscopica, un principio similare a quello utilizzato dal sistema visivo umano per percepire la profondità nell'ambiente circostante. La fotocamera è dotata di due sensori di profondità situati a una distanza fissa l'uno dall'altro (baseline di 95mm), che catturano immagini stereoscopiche dell'ambiente da due angolazioni diverse. Le due immagini acquisite presentano una certa disparità a causa della differente posizione dei due sensori.Un algoritmo di matching stereo identifica i pixel corrispondenti tra le due immagini e calcola la disparità, ossia la differenza posizionale, di ciascun pixel.Utilizzando la disparità e la baseline, l'algoritmo triangola la posizione tridimensionale di ogni pixel, creando una mappa dettagliata di profondità.

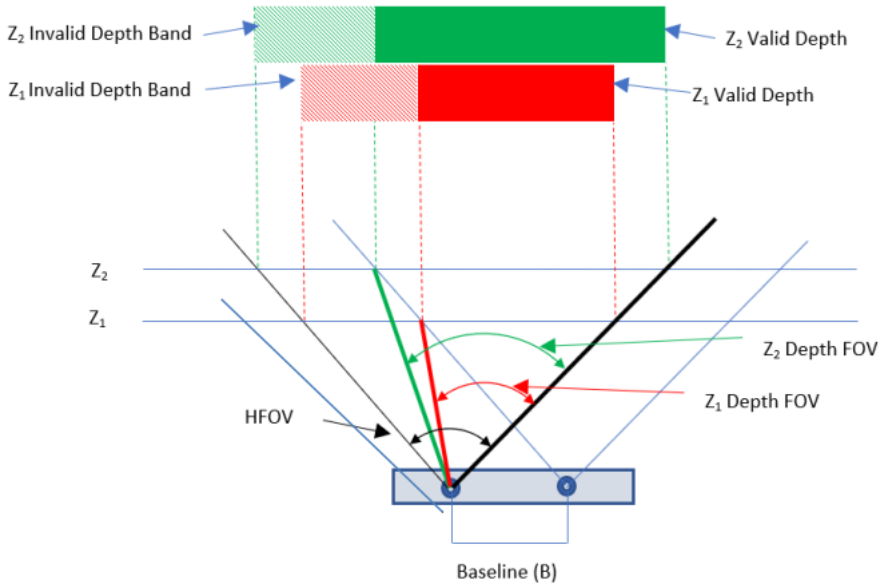


Figura 3.2: Depth

Nell'ambito delle tecnologie di imaging tridimensionale, la precisione e l'affidabilità dei dati acquisiti sono di fondamentale importanza. La fotocamera Intel RealSense D455, sebbene sia un dispositivo avanzato, non è immune da sfide comuni a tutte le fotocamere di profondità, tra cui la presenza di una '**Invalid Depth Band**'(Figura 3.3). Questo termine designa una zona specifica nell'immagine di profondità acquisita in cui i dati risultano non affidabili o inutilizzabili per analisi e applicazioni successive.

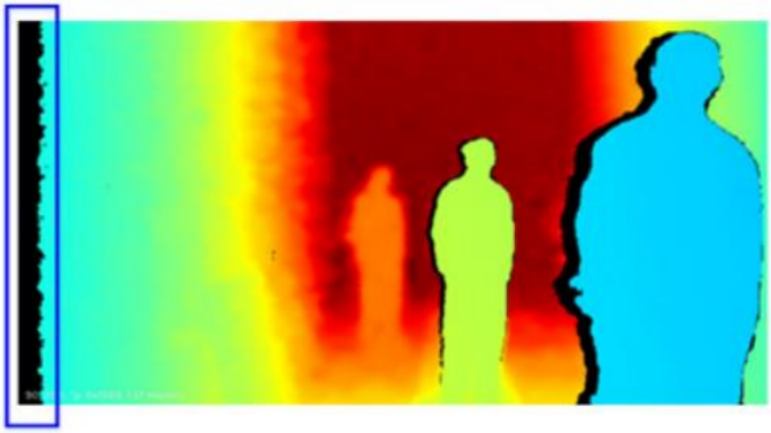


Figura 3.3: Invalid Depth Band

La **Intel RealSense D455** presenta una serie di **vantaggi** significativi che la rendono un'opzione privilegiata per diverse applicazioni nel campo dell'imaging tridimensionale:

- *Alta Precisione:* Con una risoluzione di profondità di 1280x800 pixel, la D455 è in grado di catturare immagini dettagliate, rendendola ideale per applicazioni che richiedono misurazioni precise.
- *Ampio Campo Visivo:* Con un campo visivo orizzontale di 86 gradi e verticale di 57 gradi, il dispositivo può osservare un'ampia area, facilitando l'acquisizione di dati in vari ambienti.
- *Range Operativo Esteso:* Con un range di rilevamento della profondità che varia tra 0,4 e 6 metri, offre flessibilità per diversi scenari di utilizzo.
- *Versatilità:* La D455 supporta una molteplicità di piattaforme e linguaggi di programmazione, consentendo agli sviluppatori di integrarla in una varietà di progetti e applicazioni.

Tuttavia, come ogni tecnologia, anche la D455 presenta alcune **sfide** e **limitazioni**:

- *Sensibilità alle Condizioni di Luce:* In ambienti con illuminazione intensa o diretta, la qualità dei dati di profondità può essere compromessa.
- *Limitazioni con Superfici Riflettenti o Trasparenti:* La D455 potrebbe non rilevare accuratamente superfici che riflettono o trasmettono la luce, come vetro o specchi.
- *Prezzo:* Il costo della D455 è relativamente elevato, il che potrebbe rappresentare una barriera per piccoli studi di sviluppo.

Considerando i suddetti vantaggi e svantaggi, la scelta della Intel RealSense D455 dovrebbe essere ponderata in base alle esigenze specifiche del progetto e alle condizioni operative previste.

3.2 Acquisizioni e Dataset

Per le finalità di questo studio, le acquisizioni sono state eseguite in ambiente interno, precisamente all'interno del laboratorio universitario. La scelta di un ambiente interno è stata dettata dalla necessità di avere un controllo maggiore sulle condizioni di illuminazione e sullo sfondo, fattori che influenzano significativamente la qualità dei dati acquisiti. Il **dataset** raccolto durante le fasi di acquisizione è composto da video in cui sono presenti tre soggetti differenti, ripresi contro tre diversi sfondi o **background** per aggiungere varietà e complessità ai dati raccolti. Ciò è stato fatto per simulare diversi ambienti e condizioni che potrebbero essere riscontrati in scenari reali e pratici.I soggetti coinvolti nelle sessioni di acquisizione sono tre individui diversi.

NUMERO TOTALE DI FRAME			
	SOGGETTO N°1	SOGGETTO N°2	SOGGETTO N°3
BACKGROUND N°1	153	145	259
BACKGROUND N°2	175	164	241
BACKGROUND N°3	168	148	122

Figura 3.4: Tabella riassuntiva

La diversità dei soggetti è fondamentale per conferire al dataset una varietà sufficiente da permettere l'analisi e la validazione dei metodi proposti in questo studio. Ogni soggetto è stato coinvolto in 15 sessioni di acquisizione, e ogni sessione è stata attentamente pianificata per esplorare diversi aspetti e scenari di movimento. Ogni sessione di acquisizione ha generato un video composto, in media, da 200 frame. Questo numero di frame è stato considerato adeguato per garantire una campionatura significativa delle dinamiche del movimento dei soggetti senza generare un volume di dati eccessivo e difficilmente gestibile. Le sessioni di acquisizione sono state pianificate e condotte con precisione, ponendo attenzione a mantenere costanti le condizioni ambientali e la configurazione della fotocamera. Questo approccio ha garantito che i dati raccolti fossero coerenti e comparabili tra loro.

3.3 Algoritmo Utilizzato

L'algoritmo impiegato in questo studio è il **YOLOv8 Ultralytics Pose**, un potente e avanzato sistema di rilevamento della posa progettato e ottimizzato per immagini RGB. Questo algoritmo si basa sulla ben nota architettura YOLO (You Only Look Once), nota per la sua efficienza e accuratezza nel rilevamento di oggetti e pose in immagini e video.

3.3.1 Soft Training

Per adattare l'algoritmo alle specifiche esigenze del nostro studio, abbiamo condotto una procedura di soft training. In questa fase, l'algoritmo YOLOv8 Ultralytics Pose, che di default è configurato per riconoscere e analizzare **17 keypoints**, è stato riaddestrato per lavorare con soli **5 keypoints**. Questa modifica selettiva è stata guidata dalla volontà di focalizzarsi su informazioni cruciale e ridurre la complessità del modello. I cinque keypoints selezionati sono strategici: uno situato sulla testa, due posizionati sulle spalle e due sui fianchi. Questi punti sono stati scelti per la loro rilevanza nella rappresentazione schematica e nell'analisi del movimento umano, permettendo al modello di concentrarsi su caratteristiche fondamentali dell'anatomia umana durante il processo di addestramento e inferenza. Tale approccio ha reso il processo di addestramento e inferenza più efficiente e snello, mantenendo al contempo un livello soddisfacente di accuratezza e rilevanza nell'analisi dei dati di posa. Il soft training è stato eseguito fornendo al modello immagini Depth (con applicata la colormap) associate alle corrispondenti label, contenenti informazioni dettagliate relative ai soggetti presenti nelle immagini. Le label forniscono dati preziosi che guidano e informano il modello durante il processo di addestramento, permettendogli di apprendere efficacemente le caratteristiche salienti dei soggetti e dei keypoints di interesse.

3.3.2 Primo Ciclo di Training

Nel primo ciclo di addestramento, il modello è stato addestrato utilizzando i dati associati ai background 1 e 2. Dopo la fase di addestramento, il modello addestrato è stato poi testato sul subset di dati associati al background 3. Questo approccio ha permesso di valutare le prestazioni e l'accuratezza del modello in un contesto differente da quello in cui è stato addestrato.

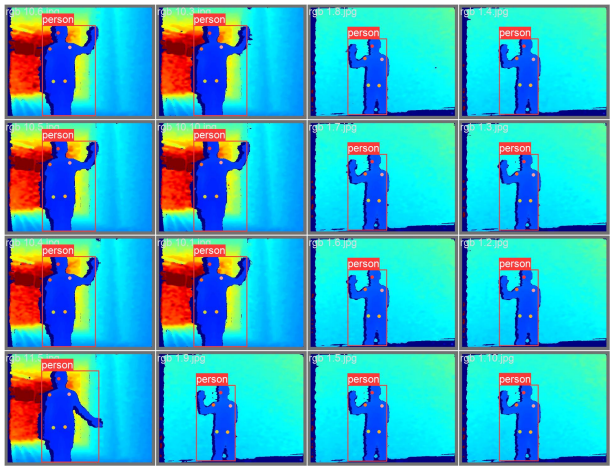


Figura 3.5: 1° Train

3.3.3 Secondo Ciclo di Training

Il secondo ciclo ha visto il modello addestrato sui dati relativi ai background 2 e 3, con la fase di testing condotta sul subset associato al background 1. Ancora una volta, l'obiettivo era osservare la resilienza e l'accuratezza del modello in un ambiente diverso, valutando la sua efficienza e affidabilità nel rilevare e analizzare i keypoints di interesse.

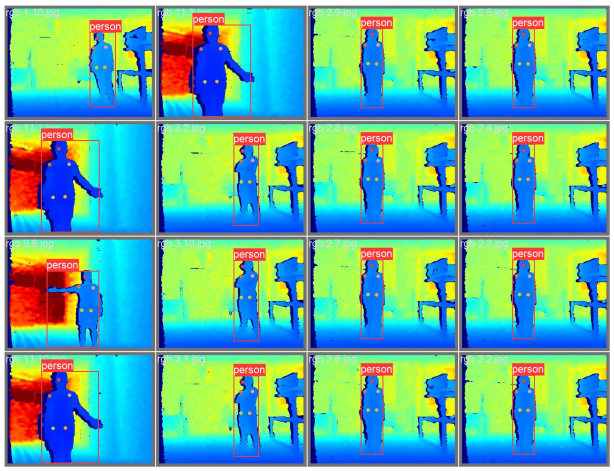


Figura 3.6: 2° Train

3.3.4 Terzo Ciclo di Training

Nel terzo e ultimo ciclo, il modello è stato addestrato combinando i dati dei background 1 e 3. Il testing è stato successivamente eseguito sul dataset associato al background 2. Questo ulteriore ciclo ha fornito ulteriori dati e insight sul comportamento del modello, consolidando i risultati ottenuti nei precedenti cicli di addestramento e testing e fornendo una visione complessiva delle capacità e delle prestazioni dell'algoritmo YOLOv8 Ultralytics Pose.

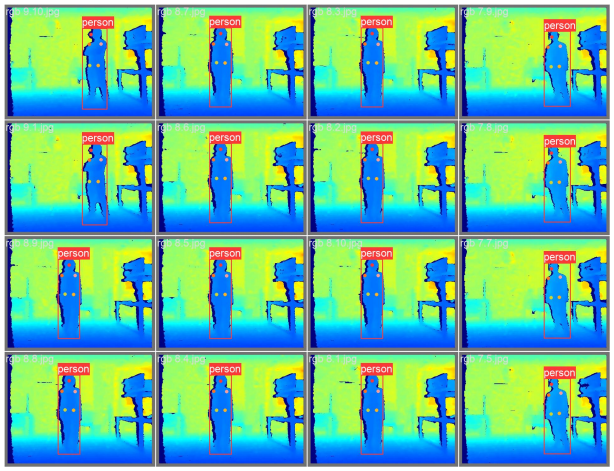


Figura 3.7: 3° Train

Capitolo 4

Risultati e Discussione

4.1 Risultati

In questo capitolo, presenteremo i risultati ottenuti dai tre cicli di addestramento dell'algoritmo e discuteremo le prestazioni del modello. Ogni ciclo di addestramento è stato seguito da una fase di testing su dati associati a background diversi, consentendoci di valutare la capacità del modello di adattarsi a contesti ambientali variabili. Durante la valutazione dei risultati, verranno forniti i valori delle metriche chiave, tra cui **Recall**, **Precisione** e **F1-Score**, ottenuti da ciascun test. Ecco una breve spiegazione di queste metriche:

- **Recall (Recupero o Sensibilità)**: Il recall misura la capacità del modello di identificare correttamente tutti i casi positivi. In altre parole, ci indica quanti dei casi effettivamente positivi il modello è riuscito a catturare.
- **Precisione**: La precisione misura la capacità del modello di classificare correttamente i casi positivi tra tutte le previsioni positive. Indica quanto sia preciso il modello nell'identificare i casi positivi.
- **F1-Score**: L'F1-Score è una metrica che bilancia il trade-off tra il recall e la precisione. È particolarmente utile quando c'è la necessità di trovare un equilibrio tra queste due metriche. Un valore più alto di F1-Score indica un buon bilanciamento tra recall e precisione.

Queste metriche sono essenziali per comprendere l'efficacia del modello nelle diverse situazioni e aiutano a valutare quanto il modello sia in grado di identificare correttamente i casi positivi e di evitare falsi positivi. Nel corso di questo capitolo, analizzeremo i risultati delle previsioni effettuate dal modello addestrato nei tre cicli, confrontandoli con i dati di posa effettivamente acquisiti, al fine di valutare le prestazioni complessive del nostro algoritmo.

4.1.1 Primo Ciclo di Addestramento e Testing

In questa sezione, presenteremo i risultati ottenuti dal primo ciclo di addestramento, che ha coinvolto il Background 1 e il Background 2. Inoltre la fase di test è stata effettuata utilizzando le acquisizioni del background 3.L'output ottenuto è il seguente:

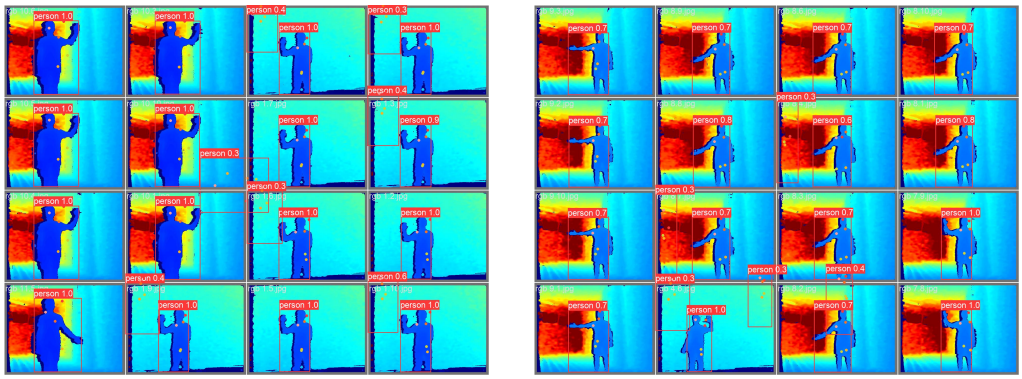
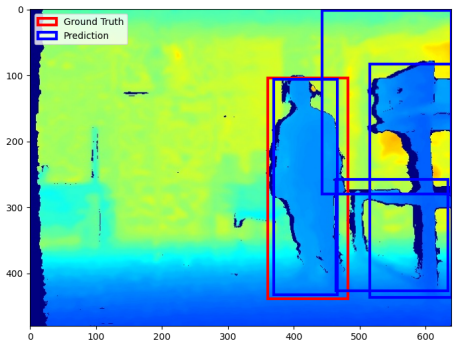


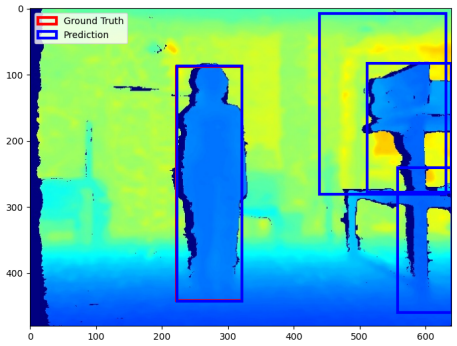
Figura 4.1: 1° Train

Nella prima fase di **testing**, sono stati eseguiti 4 test separati utilizzando 4 frame del Background 3. Per ogni test, è stata impostata una confidence di 0.6 per le previsioni del modello. I risultati ottenuti sono i seguenti:



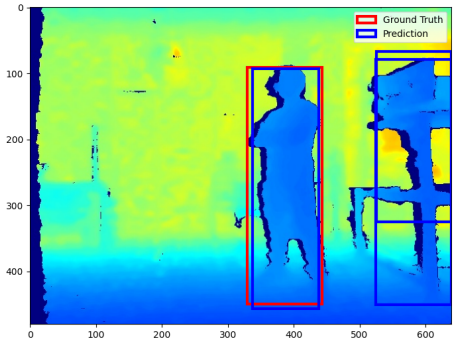
Recall	1.00
Precision	0.25
F1-Score	0.40
True Positive	1
False Positive	3

Figura 4.2: Test N°1



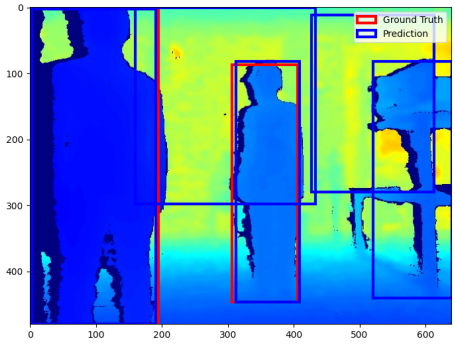
Recall	1.00
Precision	0.25
F1-Score	0.40
True Positive	1
False Positive	3

Figura 4.3: Test N°2



Recall	1.00
Precision	0.33
F1-Score	0.50
True Positive	1
False Positive	2

Figura 4.4: Test N°3



Recall	1.00
Precision	0.49
F1-Score	0.57
True Positive	2
False Positive	4

Figura 4.5: Test N°4

4.1.2 Secondo Ciclo di Addestramento e Testing

In questa sezione, presenteremo i risultati ottenuti dal secondo ciclo di addestramento, che ha coinvolto il Background 2 e il Background 3. Inoltre la fase di test è stata effettuata utilizzando le acquisizioni del background 1.L'output ottenuto è il seguente:

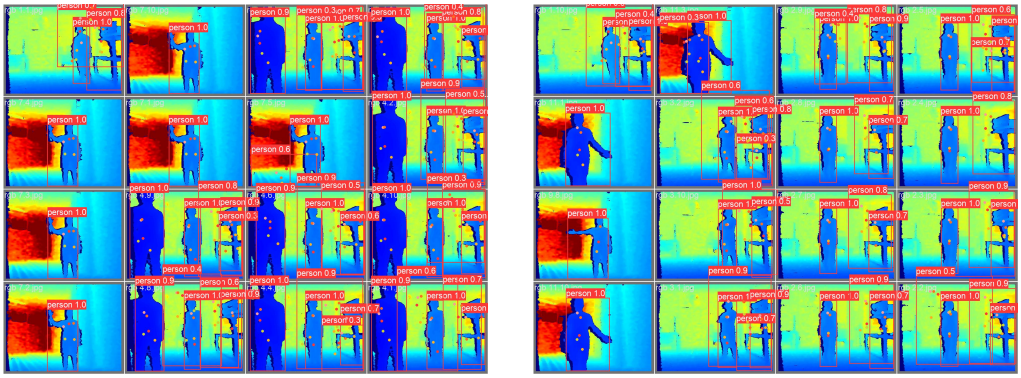
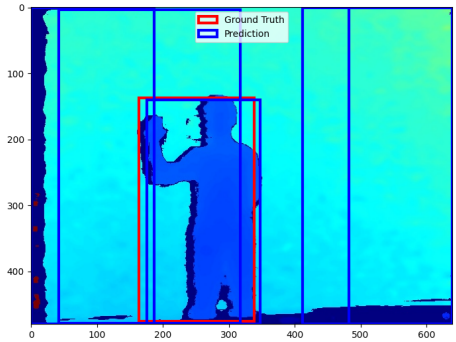


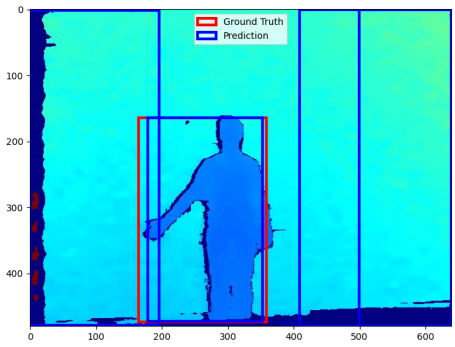
Figura 4.6: 2° Train

Nella seconda fase di **testing**, sono stati eseguiti 5 test separati utilizzando 5 frame del Background 1. Per ogni test, è stata impostata una confidence di 0.6 per le previsioni del modello. I risultati ottenuti sono i seguenti:



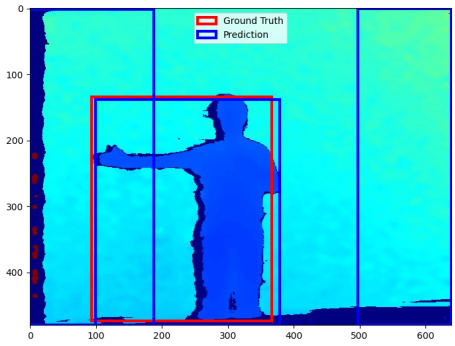
Recall	1.00
Precision	0.20
F1-Score	0.33
True Positive	1
False Positive	4

Figura 4.7: Test N°1



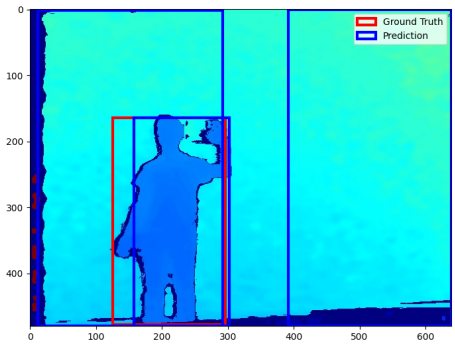
Recall	1.00
Precision	0.25
F1-Score	0.33
True Positive	1
False Positive	3

Figura 4.8: Test N°2



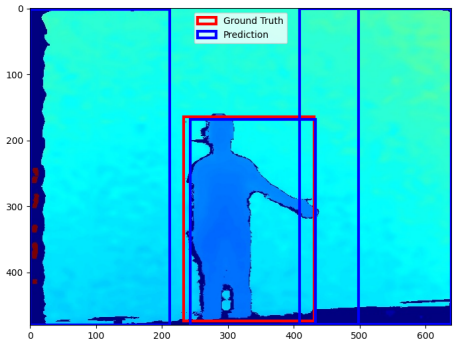
Recall	1.00
Precision	0.33
F1-Score	0.50
True Positive	1
False Positive	2

Figura 4.9: Test N°3



Recall	1.00
Precision	0.33
F1-Score	0.50
True Positive	1
False Positive	2

Figura 4.10: Test N°4



Recall	1.00
Precision	0.25
F1-Score	0.40
True Positive	1
False Positive	3

Figura 4.11: Test N°4

4.1.3 Terzo Ciclo di Addestramento e Testing

In questa sezione, presenteremo i risultati ottenuti dal ultimo ciclo di addestramento, che ha coinvolto il Background 1 e il Background 3. Inoltre la fase di test è stata effettuata utilizzando le acquisizioni del background 2.L'output ottenuto è il seguente:

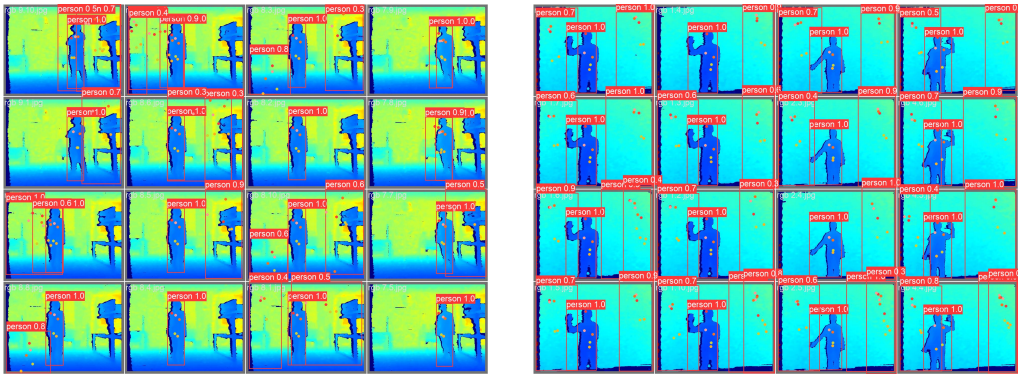
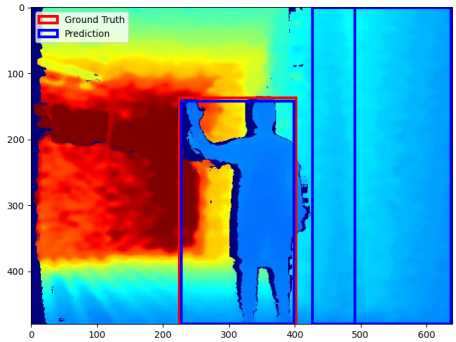


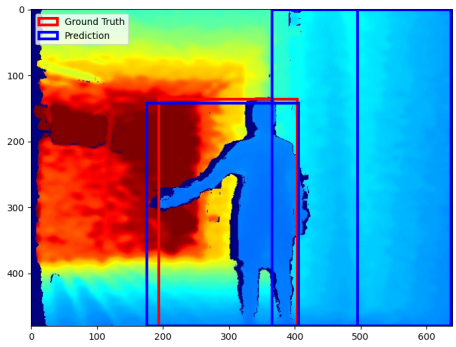
Figura 4.12: 3° Train

Nella ultima fase di **testing**, sono stati eseguiti 3 test separati utilizzando 3 frame del Background 2. Per ogni test, è stata impostata una confidence di 0.6 per le previsioni del modello. I risultati ottenuti sono i seguenti:



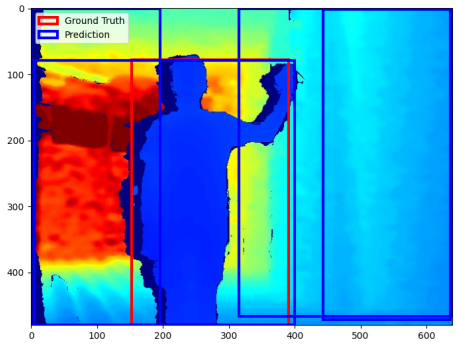
Recall	1.00
Precision	0.33
F1-Score	0.50
True Positive	1
False Positive	2

Figura 4.13: Test N°1



Recall	1.00
Precision	0.33
F1-Score	0.50
True Positive	1
False Positive	2

Figura 4.14: Test N°2



Recall	1.00
Precision	0.25
F1-Score	0.40
True Positive	1
False Positive	3

Figura 4.15: Test N°3

4.2 Commento dei Risultati

Nel corso dell'analisi dei risultati ottenuti dai tre cicli di addestramento dell'algoritmo, emerge una comprensione approfondita delle prestazioni del modello. È importante notare che la fase di previsione effettuata durante il training dell'algoritmo non è sempre precisa, e talvolta il modello tende a commettere errori. Una caratteristica comune osservata nella maggior parte dei test è l'alto valore di **Recall**, spesso pari a 1. Questo indica che il modello è generalmente in grado di identificare correttamente la stragrande maggioranza dei casi positivi durante le previsioni. Tuttavia, è fondamentale notare che la precisione delle previsioni non è altrettanto elevata. I valori massimi di **Precisione** registrati sono stati intorno al 0.40, il che significa che il modello ha tendenza a generare un numero significativo di falsi positivi. Inoltre, i valori di **F1-Score** oscillano tra 0.33 e 0.50, suggerendo un buon equilibrio tra recall e precisione, ma indicando anche che ci sono margini di miglioramento. Questo è cruciale quando si considera l'equilibrio

tra l'identificazione accurata dei casi positivi e il controllo dei falsi positivi. In sintesi, i risultati riflettono una buona capacità del modello di individuare casi positivi, ma anche una certa tendenza a generare falsi positivi. Questi risultati possono essere utili per una valutazione critica delle prestazioni dell'algoritmo e per orientare i futuri sviluppi e miglioramenti.

4.3 Direzioni Future e Miglioramenti

Alla luce dei risultati ottenuti e delle considerazioni effettuate, sono state identificate diverse direzioni future e possibili miglioramenti per il nostro progetto:

1. **Aumento del Dataset:** Una delle prime priorità dovrebbe essere l'espansione del dataset di addestramento. Questo potrebbe includere la raccolta di dati provenienti da una varietà di fonti e contesti. È particolarmente importante includere acquisizioni con sfondi diversificati e complessi, in modo che il modello possa essere più robusto e in grado di adattarsi a contesti ambientali variabili. Inoltre, l'aumento del dataset aiuta a migliorare la capacità del modello di generalizzare e ridurre il rischio di overfitting.

2. **Raccolta di Dati con Background Vuoti:** Per affrontare la sfida dei falsi positivi, potremmo raccogliere dati contenenti sfondi vuoti o uniformi. Questo consentirebbe al modello di imparare a distinguere meglio gli oggetti di interesse da sfondi privi di elementi distrattivi, riducendo così il numero di falsi positivi.

3. **Ottimizzazione dei Parametri del Modello:** Esplorare e ottimizzare i parametri del modello è un passo cruciale per migliorare le prestazioni complessive. Questo potrebbe includere l'analisi dell'architettura della rete neurale, l'ottimizzazione della funzione di costo, la regolazione dei tassi di apprendimento e l'impiego di tecniche di regolarizzazione.

4. **Training in Più Epoche:** Considerare l'addestramento del modello in più epoche. L'addestramento prolungato può consentire al modello di apprendere da più passaggi sui dati, migliorando così la sua capacità di generalizzazione.

5. **Esplorazione di Altre Metriche:** Oltre alle metriche standard come Recall, Precisione e F1-Score, è utile esaminare altre metriche specifiche per l'applicazione o il problema. Queste metriche aggiuntive possono fornire una comprensione più completa delle prestazioni del modello in scenari specifici.

Queste direzioni future rappresentano un piano per il miglioramento continuo del nostro modello. La ricerca costante, l'ottimizzazione e l'adattamento sono fondamentali per raggiungere il massimo potenziale del sistema.

Capitolo 5

Conclusioni

Il percorso di ricerca condotto nel corso di questa tesi ha aperto nuove prospettive e ha fornito una chiara comprensione delle potenzialità e delle sfide nel campo del riconoscimento delle persone mediante dati di profondità acquisiti dalla fotocamera RGB-D Intel RealSense D455. L'obiettivo principale di questo studio era esplorare e valutare la capacità di algoritmi precedentemente addestrati su immagini RGB, in particolare YOLO, nell'interpretare e processare dati di profondità.
Le conclusioni principali che emergono da questo studio sono le seguenti:

- **Generalizzabilità degli Algoritmi RGB:** Uno dei risultati sorprendenti di questa ricerca è stato l'adattamento efficace degli algoritmi originariamente concepiti per l'elaborazione di immagini RGB al contesto dei dati di profondità. Nonostante questi algoritmi non siano stati addestrati specificamente per dati di profondità, hanno dimostrato una notevole versatilità nell'affrontare questa sfida.

- **Accuratezza nei Contesti Variabili:** La nostra indagine ha incluso dati acquisiti in ambienti con sfondi diversificati, un contesto che rappresenta una sfida significativa per il riconoscimento delle persone. Tuttavia, si è notato che in alcune situazioni, possono verificarsi falsi positivi, il che suggerisce la necessità di ulteriori miglioramenti.

In conclusione, questo percorso di ricerca ha fornito un'analisi approfondita e un'accurata valutazione delle prestazioni degli algoritmi di riconoscimento delle persone basati su dati di profondità. I risultati positivi dimostrano il potenziale promettente di questo algoritmo in applicazioni reali, persino in contesti con sfondi complessi. Tuttavia, rimane ancora spazio per ulteriori sviluppi.Le conclusioni qui presentate costituiscono un solido punto di partenza per futuri studi e ricerche nel campo del riconoscimento delle persone basato su dati di profondità.

Bibliografia e Sitografia

[1] H. Zhang, D. Gračanin e M. Eltoweissy, «Classification of human posture with RGBD camera: is deep learning necessary?» In *International Conference on Human-Computer Interaction*, Springer, 2020, pp. 595–607.

[2] G. Developers. «Mediapipe Face Detection - Google Developers.» (2023), indirizzo: https://developers.google.com/mediapipe/solutions/vision/face_detector.

[3] chinmaykumar06. «face-detection-yolov3-keras.» 27/09/2023. (on Feb 24, 2021), indirizzo: https://github.com/chinmaykumar06/face-detection-yolov3-keras.

[4] «Face Detection using Cascade Classifier using OpenCV (Python).» 27/09/2023. (2023), indirizzo: https://www.geeksforgeeks.org/face-detection-using-cascade-classifier-using-opencv-python/.

[5] M. Takahashi, Y. Ji, K. Umeda e A. Moro, «Expandable YOLO: 3D Object Detection from RGB-D Images,» in *2020 21st International Conference on Research and Education in Mechatronics (REM)*, 2020, pp. 1–5. doi: 10.1109/REM49740.2020.9313886.

[6] K. S. Arikumar, A. Deepak Kumar, T. R. Gadekallu, S. B. Prathiba e K. Tamilarasi, «Real-Time 3D Object Detection and Classification in Autonomous Driving Environment Using 3D LiDAR and Camera Sensors,» *Electronics*, vol. 11, n. 24, 2022, issn: 2079-9292. doi: 10.3390/electronics11244203. indirizzo: https://www.mdpi.com/2079-9292/11/24/4203.

[7] A. Geiger, P. Lenz, C. Stiller e R. Urtasun, «Vision meets robotics: The kitti dataset,» *The International Journal of Robotics Research*, vol. 32, n. 11, pp. 1231–1237, 2013.

[8] T. Linder, N. Vaskevicius, R. Schirmer e K. O. Arras, «Cross-Modal Analysis of Human Detection for Robotics: An Industrial Case Study,» in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2021, pp. 971–978. doi: 10.1109/IROS51168.2021.9636158.

[9] T. Linder, K. Y. Pfeiffer, N. Vaskevicius, R. Schirmer e K. O. Arras, «Accurate detection and 3D localization of humans using a novel YOLO-based RGB-D fusion approach and synthetic training data,» in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, 2020, pp. 1000–1006. doi: 10.1109/ICRA40945.2020.9196899.

[10] R. Zhang, Y. Yang, W. Wang, L. Zeng, J. Chen e S. Mcgrath, «An Algorithm for Obstacle Detection based on YOLO and Light Filed Camera,» dic. 2018, pp. 223–226. doi: 10.1109/ICSensT.2018.8603600.

[11] C. A. Luna, C. Losada-Gutiérrez, D. Fuentes-Jiménez e M. Mazo, «Fast heuristic method to detect people in frontal depth images,» *Expert Systems with Applications*, vol. 168, p. 114 483, 2021, issn: 0957-4174. doi: https://doi.org/10.1016/j.eswa.2020.114483. indirizzo: https://www.sciencedirect.com/science/article/pii/S0957417420311301.

[12] .

[13] Y. Fujimoto e K. Fujita, «Depth-Based Human Detection Considering Postural Diversity and Depth Missing in Office Environment,» *IEEE Access*, vol. 7, pp. 12 206–12 219, 2019. doi: 10.1109/ACCESS.2019.2892197.

[14] T. Heitzinger e M. Kampel, «IPT: A Dataset for Identity Preserved Tracking in Closed Domains,» in *2020 25th International Conference on Pattern Recognition (ICPR)*, 2021, pp. 8228–8234. doi: 10.1109/ICPR48806.2021.9412979.

[15] D. Steininger, V. Widhalm, J. Simon, A. Kriegler e C. Sulzbachner, «The Aircraft Context Dataset: Understanding and Optimizing Data Variability in Aerial Domains,» in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops*, 2021, pp. 3823–3832.

[16] T. L. Sung e H. J. Lee, «Depth edge detection using edge-preserving filter and morphological operations,» *International Journal of System Assurance Engineering and Management*, vol. 11, n. 4, pp. 812–817, 2020. doi: 10.1007/s13198-019-00881-. indirizzo: https://ideas.repec.org/a/spr/ijsaem/v11y2020i4d10.1007_s13198-019-00881-y.html.

[17] D. Scharstein, R. Szeliski e R. Zabih, «A taxonomy and evaluation of dense two-frame stereo correspondence algorithms,» in *Proceedings IEEE Workshop on Stereo and Multi-Baseline Vision (SMBV 2001)*, 2001, pp. 131–140. doi: 10.1109/SMBV.2001.988771.

[18] P. K. Nathan Silberman Derek Hoiem e R. Fergus, «Indoor Segmentation and Support Inference from RGBD Images,» in *ECCV*, 2012.

[19] H. Luo, S. Li e Q. Zhao, «Towards Silhouette-Aware Human Detection in Depth Images,» in *2021 International Joint Conference on Neural Networks (IJCNN)*, 2021, pp. 1–8. doi: 10.1109/IJCNN52387.2021.9534347.

[20] «Dataset Towards Silhouette-Aware Human Detection in Depth Images.» 10/10/2023. (2023), indirizzo: https://pan.baidu.com/s/13hpuziavBNjS8KATClpXww.

[21] P. Schneider, Y. Anisimov, R. Islam et al., «TIMomdash;A Dataset for Indoor Building Monitoring with a Time-of-Flight Camera,» *Sensors*, vol. 22, n. 11, 2022, issn: 1424-8220. doi: 10.3390/s22113992. indirizzo: https://www.mdpi.com/1424-8220/22/11/3992.

[22] J. S. Katrolia, A. El-Sherif, H. Feld, B. Mirbach, J. R. Rambach e D. Stricker, «TICaM: A Time-of-flight In-car Cabin Monitoring Dataset,» in *32nd British Machine Vision Conference 2021, BMVC 2021, Online, November 22-25, 2021*, BMVA Press, 2021, p. 277. indirizzo: https://www.bmvc2021-virtualconference.com/assets/papers/0701.pdf.

[23] I. RealSense. «Intel RealSense D455 Depth Camera.» Accessed on 30 settembre 2023. (2023), indirizzo: https://www.intelrealsense.com/depth-camera-d455/.

Ringraziamenti

Ringrazio inanzitutto il prof. Gambi Ennio,Relatore, per avermi dato la possibilità di svolgere e portare a termine il lavoro di tesi. Ringrazio anche la prof.ssa Senigagliesi Linda, Correlatore, che mi ha aiutati durante la stesura di quest'utltima. Ho piacere di menzionare anche il dott. Nocera Antonio,che mi ha affiancato con pazienza e dedizione durante questo percorso. Le sue conoscenze e il suo aiuto pratico hanno reso questa esperienza molto più significativa e gratificante.

Un ringraziamento speciale va ai miei genitori, *Michele* e *Giovanna*,i miei primi sostenitori, coloro che hanno sempre creduto in me, e che grazie alla loro costante fiducia e supporto hanno reso possibile il raggiungimento di questo traguardo, senza il vostro sostegno non sarei mai potuto arrivare fin qui. Grazie per esserci sempre stati.

Ringrazio anche i miei fratelli, *Angelo* e *Ilaria*, le mie guide, coloro che mi hanno sempre coccolato, protetto e indicato la strada migliore da percorrere.

Un grazie lo devo sia ai miei amici di corso, in particolare a Vincenzo ed Enes con i quali ho condiviso feste, ansie, giornate intere a studiare e sia a tutti coloro che nel corso di questi tre anni hanno fatto parte di tutto ciò.

Per ultimo, ma non meno importante, vorrei ringraziare *Sofia*, colei che da più di due anni mi supporta in ogni momento della mia vita. La sua presenza, il suo amore e il suo incoraggiamento sono stati una fonte di ispirazione costante e non avrei potuto chiedere di meglio.

Grazie di cuore a tutti coloro che hanno contribuito a rendere questo percorso indimenticabile.

~Luigi