IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

ML Powered Crowd Management, Crime Prevention and Work Monitoring

M. Deepthi
Information Technology GVPCEW
Visakhapatnam, India

Vanamali Sirisha
Information Technology GVPCEW
Visakhapatnam, India

Kalamraju Niharika
Information Technology GVPCEW
Visakhapatnam, India

Prasanthi Adidela
Information Technology GVPCEW
Visakhapatnam, India

Gadam Bhavana Sai
Information Technology GVPCEW
Visakhapatnam, India

ABSTRACT:

In the constantly evolving realm of urban security, the incorporation of machine learning (ML) technologies into existing Closed-Circuit Television (CCTV) networks emerges as a game-changing solution. Inspired by real- world challenges where traditional surveillance methods often falter, our project sets out to redefine urban safety through innovative technological interventions. Our approach harness the potential of ML, involving the YOLOv8 (YOU ONLY LOOK ONCE) algorithm, to conduct real-time analysis of video feeds, swiftly identifying suspicious activities, anomalies, crowd congestion, weapons, and known criminal faces. By seamlessly integrating ML- generated insights with conventional surveillance techniques, we aim to elevate the public safety, enhance law enforcement efficiency, prevent crimes, and streamline operational processes. Additionally, our system leverages facial recognition capabilities to identify wanted individuals from a database of known criminals, further crime prevention efforts. Moreover, we incorporated a work monitoring module, for tracking employee presence, work utilization, and non-violent scenarios to optimize workplace safety. This initiative marks a significant milestone in urban security, laying the groundwork for safer and smarter cities in the future.

Key Words: Crowd Management, Violence detection, Weapon detection and Work Monitoring, sound alert and email alert.

I. INTRODUCTION

In the dynamic landscape of urban security, this project stands at the forefront, leveraging the power of machine learning (ML) technologies to revolutionize existing Closed-Circuit Television (CCTV) networks. By leveraging the state-of-the-art algorithms like YOLOv8 (You Only Look Once), the project focuses on three critical aspects: crowd management, crime prevention, and work monitoring.

The crowd management module enables real-time video analysis and crowd density estimation, allowing for the identification of overcrowding situations and facilitating efficient deployment of crowd control measures.

The crime prevention module employs ML algorithms to detects violent behaviour and non-violent behaviour, and identifies potentially dangerous weapons such as pistols and knifes and integrates a criminal face recognition to identify wanted individuals from a database of known offenders.

The Work monitoring module optimizes workplace safety by tracking employee presence, workplace utilization, and non-violent scenarios.

By integrating ML techniques with existing CCTV networks, this initiative redefines urban security and sets a new standard for proactive crowd control, crime prevention, and operational efficiency.

II. RELATED WORK

The integration of Machine learning (ML) techniques with Closed-Circuit Television (CCTV) networks has gained significant attention in recent years, with numerous studies exploring the potential of ML algorithms for enhancing urban security. Among these studies, the yolo (You Only Look Once) algorithm has emerged as a promising solution for real-time video analysis and object detection.

In the context of urban security, YOLO has been used for various applications, including crowd management, crime prevention and work monitoring. For instance, a study Wang et al (2022) integrated yolo with criminal face recognition to detect and identify wanted individuals from a database of known offenders, further enhancing crime prevention efforts.

In addition to crowd management and crime prevention, YOLO has been used for work monitoring applications. For instance, a study by Li et al. (2022) used YOLO to track employee presence and workplace utilization and workplace safety.

RetinaFace, developed by Deng et al. (2019), has gained prominence for its robustness in detecting faces with high accuracy, especially in challenging conditions such as occlusion and variations in pose.

RetinaFace also offer face alignment capabilities, aligning detected faces to a standardized pose, which is crucial for subsequent face recognition tasks. ResNet50, a deep convolutional neural network architecture developed by He et al. (2016), has been widely adopted for face recognition tasks due to its exceptional performance in feature extraction and classification.

III. EXISTING SYSTEM

The current surveillance system relies on YOLOv4 for real-time object detection within CCTV networks, but it has limitations including dependency on predefined class, reduced accuracy in crowded scenes and reliance manual monitoring, leading to errors and delayed responses.

To overcome these challenges and improve urban surveillance, an enhanced system integrating advanced ML techniques in necessary. By leveraging cutting-edge algorithms like deep learning and ensemble methods, the system can adopt to emerge threats, handle complex scenarios like crowd density estimation, and automate monitoring processes for quicker detection and response to security incidents, ultimately enhancing public safety

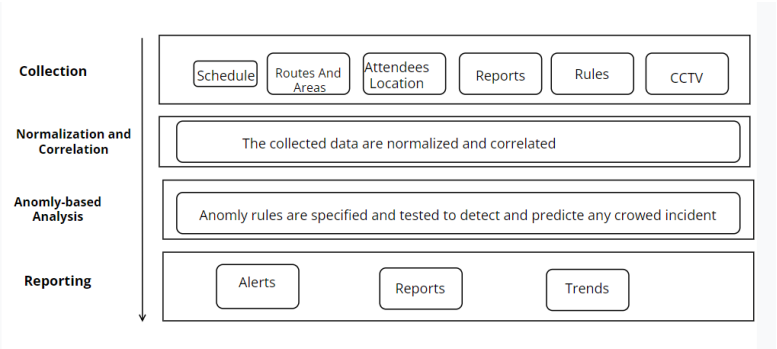


Fig.1. Existing System

IV. PROPOSED SYSTEM

The proposed system, a smart Surveillance and Safety Enhancement System (SSSES), is a comprehensive ML- based application developed using the Streamlit framework and incorporated the state-of-the-art YOLOV8 object detection model. It offers user-friendly interface for configuring tasks like crowd detection, crime prevention, and work monitoring, with support for multiple input sources including images, videos and webcams. The system can analyze input data to detect and count people, categorize crowd density levels, and provide alerts or notifications for high crowd density situations, ensuring effective crowd management. Additionally, it can identify potential violent actions or situations and presence of weapons like pistols and knives, enabling prompt responses for crime prevention. Leveraging RetinaFace for face detection and alignment, along with ResNet50 algorithm for face recognition, the system can identify individuals in real-time and cross-reference them against criminal databases for enhanced security. The system's object detection and tracking capabilities can be employed for work monitoring, aiding in compliance, hazard identification, and workflow optimization. With support for multiple input sources, the proposed system ensures compatibility with various surveillance setups and environments, making it a versatile solution for enhancing safety, security, and operational efficiency across diverse domains.

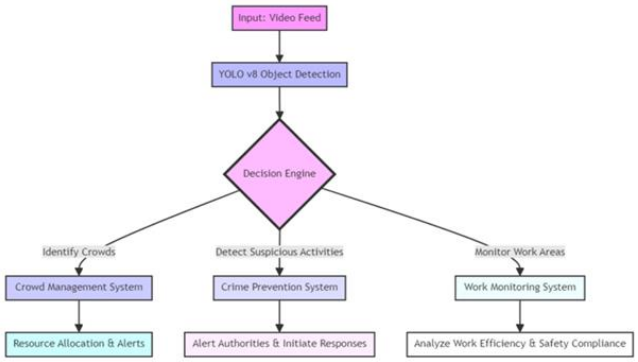


Fig.2. Proposed System

V. DATASETS USED

People_counterv0 () data set

The "people_counterv0 ()" dataset isa popular object detection dataset containing 4,574 labelled images split into train, test and validation sets. It is designed for the binary classification task of detecting people/no people in images. The dataset has two classes: "null" representing images without any "people" representing images containing one or more human figures. With its reasonably large size and clear labelling, this dataset provides a valuable model specifically aimed at identifying and localizing people in visual data, Which has applications in areas like surveillance, crowd monitoring, robotics, and scenario requiring detection and tracking of human subjects from cameras or image sources.

Violence-weapon-detection

The "violence-weapon-detection" dataset is designed for training machine learning models to detect violent scenarios and the presence of weapons in images or video frames. It contains 5,953 images related to guns, knives, violent and non-violent scenarios, split into a train set of 4583 images and a valid set of 948 images and a test set of 422 images. It applies various data augmentations and provided in multiple formats including YOLO, COCO JSON, and Pascal VOC.

Pistol and knife dataset

The "pistol and knife" dataset designed for training machine learning models to detect pistols and knives. It contains 4208 images related to knives, pistols and person scenarios split into a train set of 4064 images and valid set of 144 images with the latest v2 version released on March 16, 2024 and it applies various data augmentations and provides download formats like YOLO, COCO JSON and Pascal VOC.

Custom dataset

Custom dataset for criminal face recognition comprising images of notorious criminals organized into folders named after the individuals. Each folder contains multiple images of the respective criminal, enabling models to learn varied appearances and factors like aging, hairstyle changes etc. This dataset can prove valuable for law enforcement, enhancing public safety efforts by enabling accurate identification and tracking of individuals with criminal backgrounds.

Workplace_monitoring dataset

The workplace monitoring dataset contains 174 images depicting workplace postures/activities like standing, sitting, bending, squatting and lying, divided into a train set of 121 images, a valid set of 34 images, and a test set of 19 images, with the current v3 version released on June 8, 2023. It does not apply any data augmentation but offers download formats like YOLO, COCO JSON and Pascal VOC.

VI. ALGORITHMS USED

YOU ONLY LOOK ONCE(YOLOv8):

YOLOv8 is the latest version of the YOLO (YOU ONLY LOOK ONCE) object detection algorithm, developed by Ultralytics.

The state-of-the-art YOLOv8 (YOU ONLY LOOK ONCE) object detection model is employed as the core component for various tasks including crowd management, violence detection, weapon detection and work monitoring.

In crowd management scenarios, it can detect and track individuals, estimate crowd density. For violence detection, it can recognize aggressive actions, fights or suspicious behaviour and for weapon detection, yolov8 can be trained to identify firearms, knives enhancing security measures. In work monitoring, the algorithm can track worker movements and detects potential safety violations.

It takes an input image or video frame and divides it into a grid of cells. For each cell, the algorithm predicts a set of bounding boxes and confidence scores for object present within that cell, along with the class probabilities for each bounding box. The final output is a set of bounding boxes with their corresponding class labels and confidence scores.

YOLOv8 incorporates several improvements over previous versions, such as backbone network for feature extraction, optimized anchor box, advanced data augmentation technique.

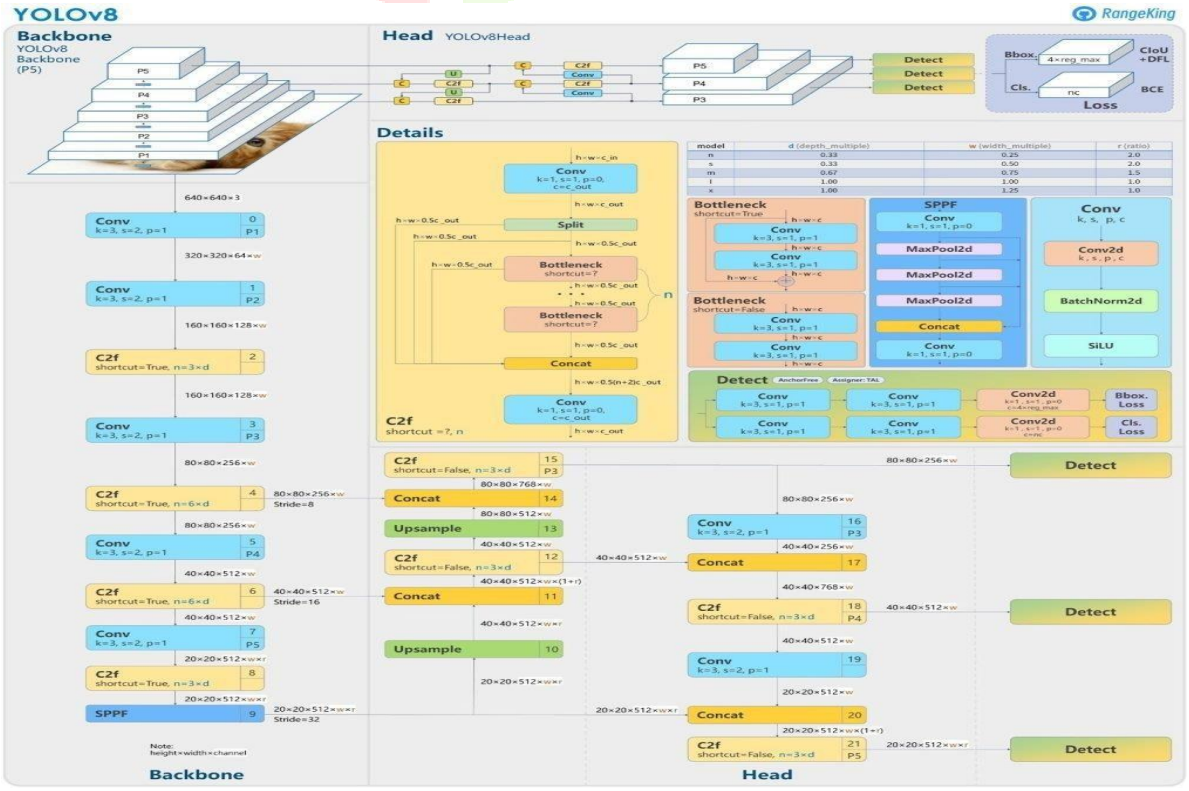


Fig.3. YOLOv8 Architecture

ResNet50:

ResNet50 is a deep convolutional neural network architecture that is widely used for various computer vision tasks including image classification, object detection and feature extraction.

During the face recognition process, the input image or video frame would passed through the ResNet50 model to extract the facial features. These features would then be compared against the trained classification model to determine if a match exists in the database of known identities.

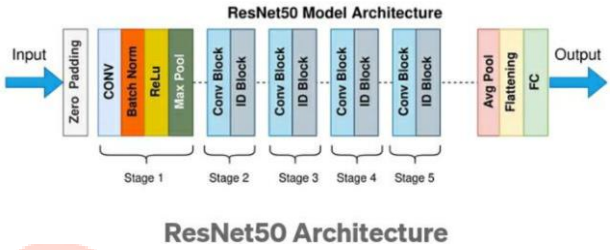


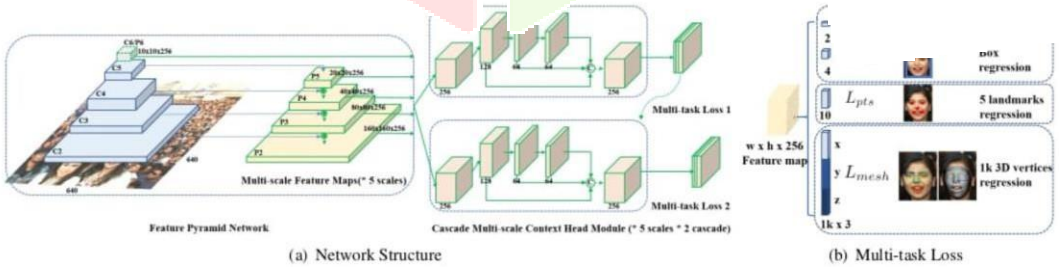
Fig.4. ResNet50 Architecture

RetinaFace:

RetinaFace is a state-of-the-art face detection and facial landmark estimation algorithm that combines robust face detection with precise facial key point localization**.**

RetinaFace can be used to accurately detect and localize faces within the input image or video frame, providing bounding boxes around each detected face. The output of Retinaface can then be used to preprocess the facial images before passing them to the face recognition component.

By combining RetinaFace for accurate face detection and alignment and ResNet50 for robust feature extraction and recognition, the crime prevention module can effectively identify individuals in real-time, enabling prompt alerts and responses to potential threats or criminal activities.



(a) Network Structure (b) Multi-task Loss

Fig.5. RetinaFace

VII. METHODOLOGY

Data Collection:

Obtained the datasets people_counterv0, violence-weapon-detection, pistol_knife and workplace_monitoring for crowd monitoring, violence detection, weapon detection and work monitoring scenarios.

Data Preparation:

Preprocessed the datasets by performing tasks such as data cleaning, data labelling, and splitting the datasets into training, validation, and testing sets for model training and performance evaluation.

Model Training

YOLOv8:

Trained the YOLOv8 model using the prepared training data, updating the model's weights through backpropagation and optimization algorithms. Monitor the training progress by tracking metrics including loss function values, precision, recall, and mean Average Precision (mAP) on the validation set. Implement techniques such as data augmentation, dropout to enhance model generalization and prevent overfitting.

RetinaFace and ResNet50:

Incorporated RetinaFace for face detection and alignment, ensuring accurate localization and alignment of faces within the images and then utilized the ResNet50 for face recognition, enabling the system to identify individuals by matching detected faces against known identities in a database**.**

Model testing and evaluation:

Performed testing on the application and its modules using both prepared datasets and real-world scenarios. Evaluated the performance of the YOLOv8 model, Retinaface, and RetNet50 for face detection and recognition based on predefined evaluation metrics.

Iteratively refined and improved the system based on the evaluation results, adjusting parameters, and fine- tuning models as necessary to optimize performance and accuracy.

Formulas for evaluation metrics:

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

$$\text{Precision: } \frac{TP}{TP+FP}$$

$$\text{F1-score: } \frac{2TP}{2TP+FP+FN}$$

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

VIII. RESULTS

The below confusion matrices and evaluation metrics showcases the performance of our yolov8 modules developed for crowd detection, violence detection, weapon detection, and work monitoring. The crowd detection module accurately identifies the crowd. The violence detection module demonstrates proficiency in recognizing non-violent and violent situations but faces challenges in precisely classifying guns and knives with an accuracy of 85.7%. The weapon detection module excels at identifying knifes, pistols, and people, with some room for improvement in distinguishing people from pistols and with an accuracy of 88.1%. The face recognition module accurately identifies the criminals in our database with an accuracy of 80%. The work monitoring module exhibits robust background and scenarios like desks, employees, and walkways, enabling efficient work monitoring with an accuracy of 89.99%.

Crowd Management

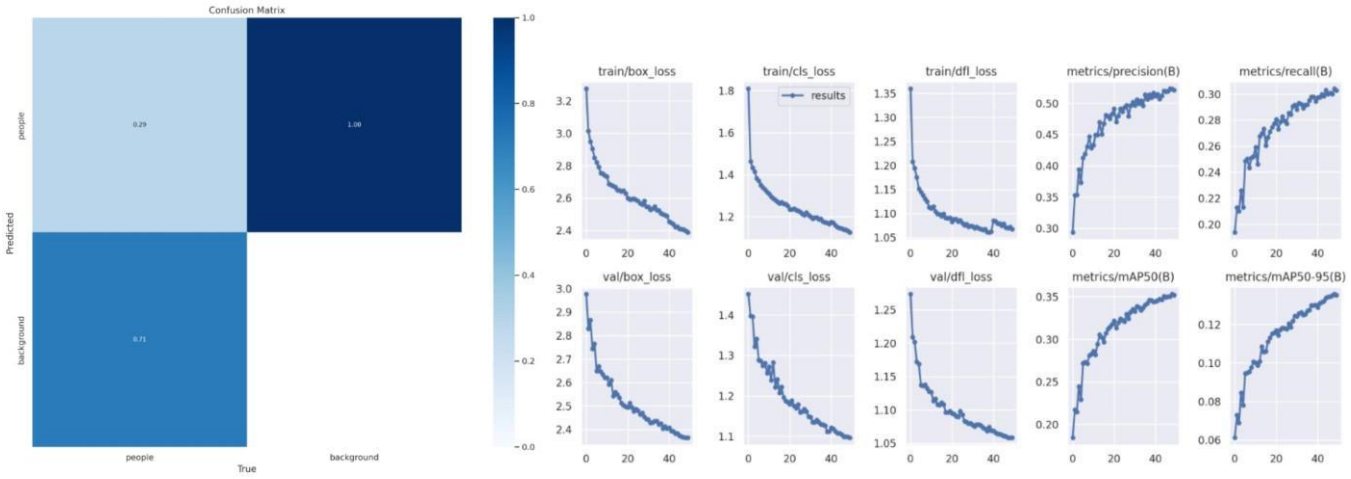


Fig.6. Confusion Matrix and evaluation metrics for crowd management

Violence Detection

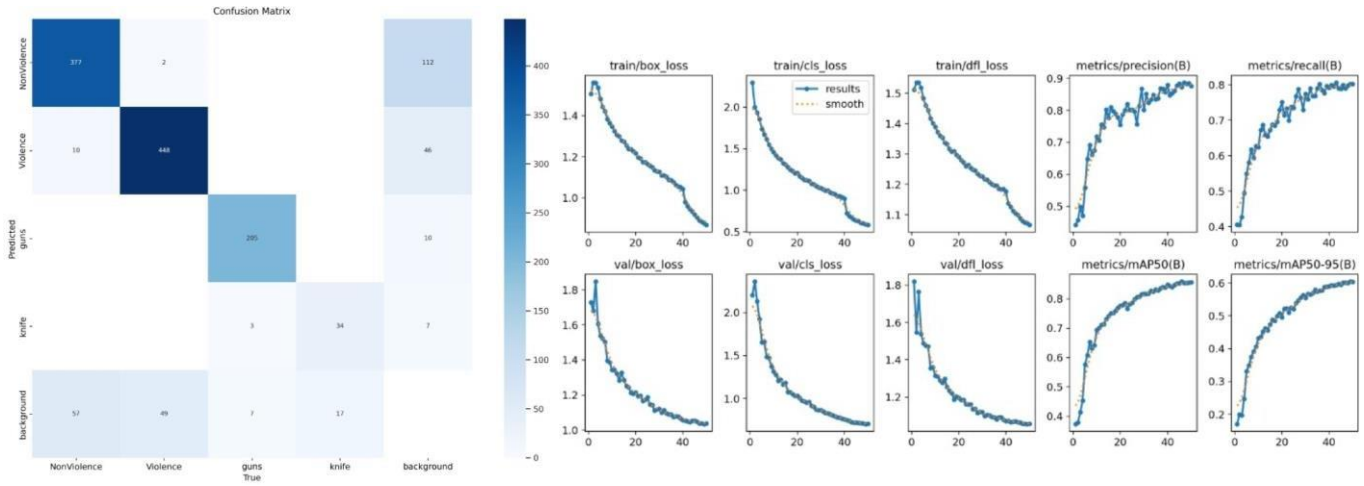


Fig.7. Confusion Matrix and Evaluation metrics for Violence detection

Weapon Detection

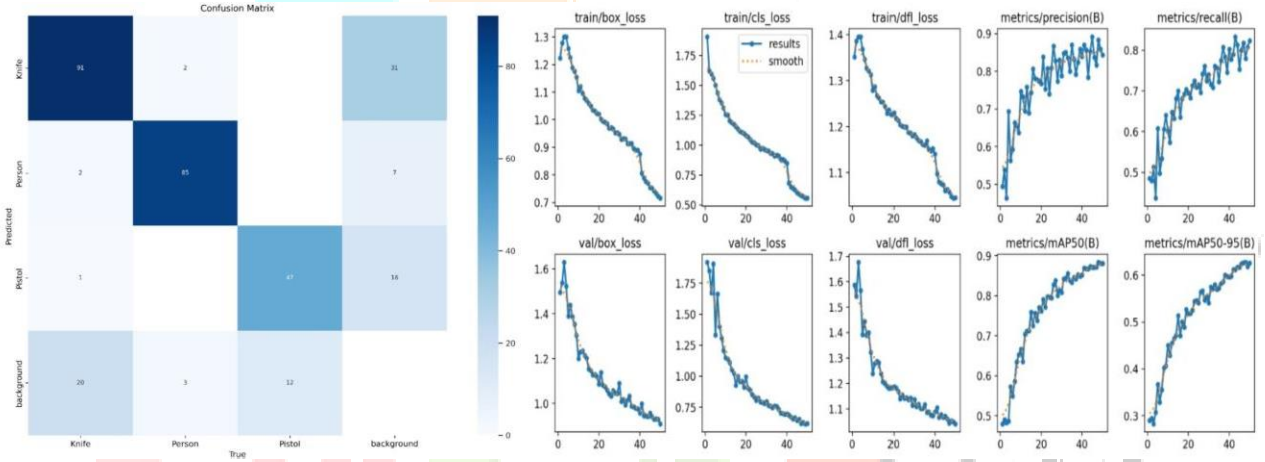


Fig.8. Confusion matrix and evaluation metrics for weapon detection

Work Monitoring

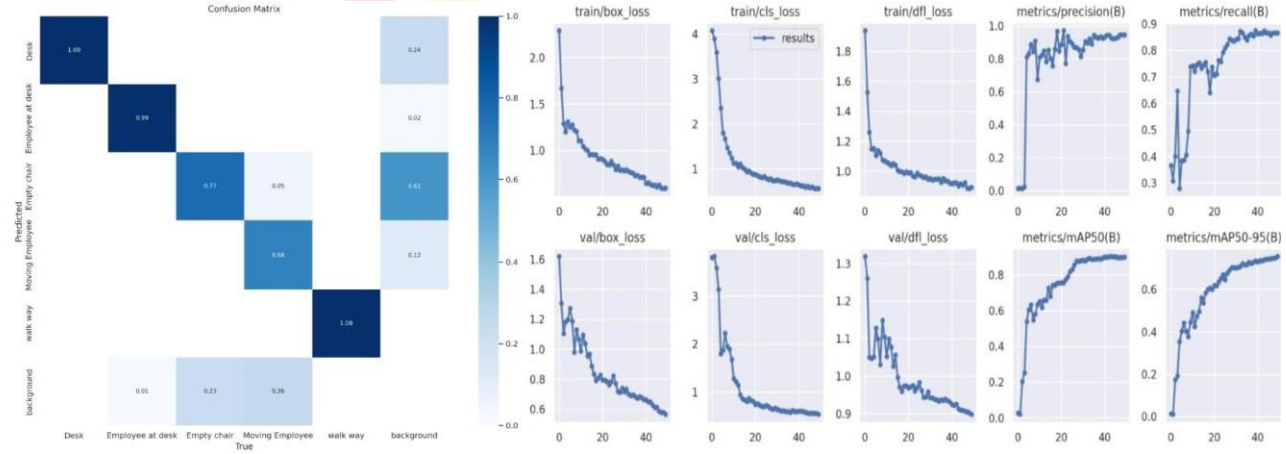


Fig.9. Confusion matrix and evaluation metrics for work monitoring

Face Recognition

Fig.10. Accuracy for Face recognition

IX. CONCLUSION

In this research, a system is presented, integrated advanced machine learning techniques with existing CCTV networks. The framework leverages YOLOv8 state-of-the-art algorithm for object detection, RetinaFace for face detection and ResNet50 for face recognition.

The Crowd Management module enables real-time video analysis, crowd density estimation, and identification of overcrowding situations, facilitating efficient deployment of crowd control measures. The crime prevention module to detect violent behaviour and identifies weapons such as pistols and knives, and integrated criminal face recognition to identify wanted individuals from a database. The work monitoring module optimizes workplace safety by tracking employee presence and work utilization.

As future work, conducting comprehensive research to identify and evaluate the most suitable models and techniques for collecting and processing multimodal data sources, including video feeds, audio signals, and sensor data enhancing the system's intelligence-based security and monitoring capabilities and exploring the integration of other sensor data, such as audio or infrared sensors, to enhance the system's ability to detect and respond to potential threats or anomalies.

X. REFERENCE

[1]W. Halboob, H. Altaheri, A. Derhab and J. Almuhtadi, "Crowd Management Intelligence Framework: Umrah Use Case," In IEEE Access, vol. 12, pp.6752-6767, 2024.

[2] S. Kilic and M. Tuceryan, "Crime Detection from Pre-crime Video Analysis with Augmented Pose Information," 2023 IEEE International Conference on Electro Information Technology (eIT), Romeoville, IL, USA, 2023, pp. 1-6.

[3] A. Goenka and K. Sitara, "Weapon Detection from Surveillance Images using Deep Learning." 2022 3rd International Conference for Emerging Technology (INCET), Belgaum, India, 2022, pp. 1-6.

[4]A. Kumari Sirivarshitha, K. Sravani, K.S. Priya and V. Bhavani, "An approach for Face Detection and Face Recognition Libraries in Python," 2023 International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 2023, pp. 1274-1278.