Journal of Computing & Biomedical Informatics
ISSN: 2710 - 1606
Volume 06 Issue 01
2023

Research Article
https://doi.org/10.56979/601/2023

A Deep Learning Based Approach to Enhance Object Edge Detection for Office Surveillance System

Samra Mushtaq[1], Kaleem Razzaq Malik[1], Zohaib Ahmad[2], and Muhammad Sajid[1*]

[1]Department of Computer Science, Air University Islamabad, Multan Campus, Multan 60000, Pakistan.
[2]Department of Criminology & Forensic Sciences, Lahore Garrison University, Lahore, 54000, Pakistan.
[*]Corresponding Author: Muhammad Sajid. Email: msajid@aumc.edu.pk

Received: August 05, 2023 **Accepted:** December 01, 2023 **Published:** December 05, 2023

Abstract: Office surveillance systems are essential for ensuring safety and monitoring activity in an office. These systems use a technique called object edge detection to track and trace objects during live action. Deep Learning is used in this study that improves object boundary detection and feature extraction in office monitoring systems. The most crucial aspect of the surveillance system is to detect and monitor human activity. As the advancement in video surveillance technology increases continuously, the fundamental task is to accurately detect objects by utilizing object detection techniques. Object detection techniques are widely used to achieve this goal by leveraging information gathered from images. By utilizing a stationary camera, the identification of humans in recorded videos can be done rapidly. This study uses the YOLOv8 model with two datasets, focusing mainly on the furniture dataset to detect table objects and changes that occur in them. Then compile and integrate datasets specific to office situations, including different illumination, object scale, and occlusion. Our proposed approach achieves excellent performance measures, including mAP50 with an accuracy of 99.3% on the Furniture Dataset. We comprehensively evaluate our approach to ensure its correctness. Also apply the YOLOV8 model to the Person Dataset to detect human activity without disclosing the individual’s identity, achieving an accuracy of 70.1%. To maintain a log of all object changes and authorized and unauthorized personnel, a database is created that contains all relevant information for security purposes.

Keywords: Office Surveillance; Object Detection; Image Segmentation; Feature Extraction; Data Overlapping; Deep Learning; YOLOv8 Model.

1. Introduction

The most crucial aspect of the surveillance system is to detect and monitor human activity. As the advancement in video surveillance technology increases continuously, the fundamental task is to accurately detect objects by utilizing object detection techniques. Object detection techniques are widely used to achieve this goal by leveraging information gathered from images. By utilizing a stationary camera, the identification of humans in recorded videos can be done rapidly. This quick detection process holds significant importance for a wide range of applications, especially for office surveillance and conference rooms [1]. To address this need, a deep learning model is used to recognize objects at multiple levels by collecting and analyzing image data.

Automatic object detection has become a hot topic in recent times with numerous approaches being proposed in the literature. Deep Learning (DL) has now taken center stage, and its automatic object detection approaches are progressively replacing earlier techniques based on classical computer vision. These newer approaches, such as DPM [2] and Selective Search [3], offer significant improvements in terms of both speed and reliability over their predecessors. With these advantages, DL-based approaches are the way to go for anyone seeking effective and efficient object detection solutions. Object detection can be a challenging task [4], and it becomes even more difficult when dealing with small objects. It is crucial to address this challenge by using advanced methods and techniques that can accurately detect and classify small objects. Failing to address this challenge can lead to errors and inaccuracies in the detection process, which can have serious consequences. Moreover, efficient and sustainable office surveillance requires innovative solutions that can overcome the challenge of detecting small objects from a distance [5-7].

Expanding the study aims and model, this study's pursuit of innovation has resulted in the creation of groundbreaking deep learning models, with the Yolov8 model leading the charge. These models have undergone meticulous training and evaluation to ensure their proficiency in processing a database of table and person images. By utilizing this advanced model, this study has achieved unparalleled accuracy and efficiency, setting a new standard in deep learning. Nowadays, it is crucial to ensure that we have robust office surveillance systems in place. That's why this research is so crucial - it's underpinned by a well-defined set of objectives that will help us achieve our goals of enhancing security, preserving privacy, and promoting workplace productivity [8]. With the implementation of deep-learning models like the YOLOv8 model, we can identify and implement more efficient methods to fortify security and privacy within the workplace. The focus on elevating employee workspace privacy will foster an environment conducive to increased productivity, while the use of cutting-edge techniques for more accurate object detection and tracking will provide much-needed peace of mind.

Lastly, unauthorized person identification through facial and hand recognition will enable us to respond rapidly to potential security threats or breaches. For detection, segmentation, and feature extraction YOLOv8 model is used in this study. Ultimately, our findings will pave the way for future explorations and technological developments in sustainable office surveillance. The research is structured logically manner, effectively presenting its findings. The first section is dedicated to highlighting the significance of the study problem, emphasizing the relevance of computer-aided design in office surveillance and the challenges faced when automatically identifying objects. The section effectively demonstrates how these obstacles were overcome, creating a foundation for the research. The rest of the paper is as follows: Section 2 provides an exclusive overview of the previous study on deep learning, object detection, and anomaly detection. Section 3 describes the proposed method for object detection. Section 4 focuses on the experiments conducted to evaluate the proposed methodology. Section 5 finally, the research concludes with a summary of the main findings and contribution along with future work.

2. Related Work

In this survey, 167 studies are based on deep learning to identify the latest trends for resource-constrained devices to detect objects. Lightweight models present supreme performance on edge devices, and the review grouped techniques used to create models, input image types, device types, and application types addressed by the models. Overall, this field has performed well in the last decade [9]. Similarly, [10] utilized a Faster R-CNN fused with FPN to detect guns on CCTV images. Evaluation of CCTV data revealed that while non-CCTV data give less accurate results on small object detection, synthetic training data give more accurate results. This highlights the value of FPN integration for improving general-purpose object detection in small objects from CCTV footage. Computer vision tech can process overhead images from several resources for various purposes. This article surveys the latest object detection techniques based on deep learning, introduces new approaches, and compares publicly available datasets to enhance research in this area [11]. YOLO [12-14] and Single Shot multi-box Detector (SSD) [15] are used as object detectors that have gained immense popularity due to their direct object localization and classification. The object detection results using this detector architecture have proved very effective. DL modules like FPN [16] have been developed to improve general-purpose object detection in CNN architectures. The latest works in CCTV use DL architectures with FPN fusion for handguns and knife detection from [10, 17]to tackle small object size issues.

Authors in [17] proposed an object detector fused with multi-level FPN to detect handguns in CCTV by using a custom dataset. For security purposes, MSD-CNN is used in a surveillance system. It is used for the classification of objects into different subclasses, locating the objects, and then detecting guns and knives. The results show 85.5% accuracy in a multi-view camera for object detection [18]. The algorithm utilizes optical flow and a knowledge-based CNN for the detection and tracking of multiple objects in surveillance video streams with high accuracy in complicated environments [19].To exclude the damaged areas, this article proposed a solution to recognize and classify objects in difficult weather conditions by using the Canny operator. This method outperforms CNNs in binary classification (79% vs. 65% accuracies) and achieves 71% to 86% accuracy in multiclass classification [20]. Fast R-CNN, a deep neural network framework, achieves high accuracy and speed in object detection, outperforming other methods with mean average precision (mAP) of 89.4-91.33%. And can detect both 2D and 3D objects [21]. YOLOv4 algorithm that is based on deep learning is proposed in this study. It is used to detect real-time objects and classification of indoor construction elements with high automation that gives a precision of 92% with 0.83 of average loss [22]. Cities face traffic congestion due to growth. YOLOv5 model detects cars, pedestrians, and traffic lights in real time. The proposed system improves vehicle recognition in various weather, reducing congestion, and increasing safety with 72.3% and 57.3% accuracy for cars and trucks respectively, and mAP (0.5) [23].

Proposed a new method for detecting checkerboard corners using YOLOX and Harris algorithm. Achieves superior accuracy and robustness against image degradation. Suitable for automatic detection and camera calibration [24]. A DL-based approach was developed in a surveillance system for the real-time detection of handguns and knives by using low-cost SBC devices. This approach achieved near real-time performance, making it affordable and effective [25]. ANPR is crucial to managing vehicles. This uses a CNN-based approach that accurately predicts corner points and performs rectification with excellent results on the CCPD dataset [26]. Proposed (Hybrid DCACN) with an Improved Whale Optimization (IWO) framework that detects and classifies road damage images with 98.815% accuracy, enabling efficient transport infrastructure management [27]. A novel scheme of combining ACDC and DAWN datasets with custom training weights improved the object detection accuracy using the YOLOv8 model under adverse weather Conditions for autonomous driving. This approach demonstrates competitive result improvement for object detection under difficult weather conditions [28].

Every human activity is specific and different from every other person, making this a difficult task to detect anomalies in the surveillance system. It is another difficult task for the reasoning of human activity detection from different datasets due to more development in deep learning models and algorithms [29]. In the smart city era, video surveillance has become essential to improve the quality of life. Video cameras for surveillance purpose are mostly installed at specific places to make sure that it covers the whole region of interest or areas. Video data-driven systems are also useful in offices, healthcare, transportation, factories, schools, shopping malls, malls, etc. Identification of suspicious activity at airports, bus stops, and train stations [30], unusual activity at public gatherings [31], and unusual patterns followed by factory workers [32]. This paper emphasizes the need for efficient anomaly detection techniques and their crucial role in ensuring the safety and well-being of citizens in smart cities and presents a survey of efficient methodologies to ensure citizen safety [33]. CCTV surveillance is often slow to respond to criminal activities. Automatic weapon detection can aid investigations and prevent criminal activities in public places.

DL techniques outperform machine learning techniques in detecting different types of weapons [34]. The primary objective of the research strategy seems to be to use Deep Learning for identifying objects such as tables and persons. This research uses the Yolov8 model to achieve this objective. By leveraging the architectural layer of the model, this can identify useful characteristics that are essential for accurately identifying the objects. From the previously mentioned literature, this study proposed a novel approach based on deep learning models for maintaining privacy and security. Validation of region of interest by recognizing and segmenting the objects. The focus of this study on the use of Deep Learning and these advanced architecture models highlights the importance of machine learning in addressing complex problems, such as office surveillance. This study significantly aids in the blooming of efficient and long-lasting security management strategies for the surveillance system.

3. Materials and Methods

To conduct a successful experiment, it was essential to follow the meticulous procedures outlined in the subsequent subsection. To begin with, the process of selecting all the pertinent datasets was discussed. The first subsection briefly explains the process of collecting and annotating the custom datasets. Then we discuss the training phase of the custom dataset along with the validation and testing phase as well as their evaluation. With these precise steps, we were able to conduct a thorough and efficient investigation.

3.1 Dataset Collection and its Annotation

This study utilized two datasets for the main contribution. The first one was the Furniture Dataset, which was predominantly employed for Object detection and was annotated for one class labeled as Table with class_id '0' and some part of this dataset goes through Image Segmentation. Polygon annotation was done on all the images that were selected from the Furniture dataset for segmentation purposes, and were also annotated for one class labeled as Table with class_id '0'. The second dataset was the Person Dataset, which was employed to Detecting persons in Office Surveillance and annotated with eight classes labeled as authorize person1, authorize person1's face, authorize person2, authorize person2's face, servant, servant's face, unauthorized person, unautherized person's face with class_id (0,1,2,3,4 and so on). Both the datasets were collected manually as the Furniture Dataset was gathered from Google by saving the images and the Person Dataset was collected by capturing images from camera video of Office Surveillance. Two tools were used for the annotation step for Detecting Objects and Image segmentation. For the manual data annotation of this work, the "makesense.ai" (https://www.makesense.ai/) and "roboflow" (https://roboflow.com/) websites were utilized because these are reputable platforms for data annotation. Both the datasets were annotated manually by using the above-mentioned tools for Object Detection and Image Segmentation purposes. Figure 1 shows the sample images for the custom datasets.

3.2 Data Preprocessing

Data processing is a crucial part of Object Detection to attain accurate results. In this work, we recognized the importance of data processing after annotating all the appropriate images from the custom datasets. To ensure the best possible outcome, we used the YOLOv8 algorithm to prepare the datasets for training, validating, and testing. Two versions of image size are taken by YOLOv8 for training such as 640 × 640 and 1280 × 1280. However, the image size of both the datasets used in this work by Yolov8 is 640 × 640. However, the augmentation process cannot be performed by the YOLOv8 model because it does not have this function. Therefore, we used the Roboflow website which helped us in the processing of the YOLOv8 method. Roboflow's website helped us with data preprocessing and augmentation techniques, including resizing, horizontal and vertical flips, rotations, noise, and more. Both datasets are split into three parts training, validating of data, and testing images with a ratio of 70%, 20%, and 10%, respectively. The splitting of both datasets can also be done by using the Roboflow website. We put a small portion of the Person Dataset in the Furniture Dataset during training and validation because the Person Dataset contains images that have tables.

First, the training, validation, and testing are done on the Furniture Dataset to detect Table then all the training, validation, and testing is done on the Person Dataset to detect persons. One more test is done on the person dataset by using weights of the Furniture Dataset to detect the Table in the Person Dataset. The image size remains 640 × 640 during all stages of training, validation, and testing of both datasets. Our methodology ensured that we had a robust and reliable dataset for object detection, and we believe that our approach is an effective and efficient way of preparing datasets for object detection.

(a) (b)

Figure 1. Sample images for the Custom datasets (a) For the Furniture Dataset and (b) For the Person Dataset

3.3 Model Training

For training and result assessment by validating and testing, we utilized Google Colab with a Python-based approach. To annotate the datasets manually, two websites are used such as "makesense.ai" and "roboflow". The YOLOv8 was the most recent version of YOLO until this study. These are the new versions based on deep learning which are much quicker than old algorithms and have a higher degree of accuracy in the field of detecting objects. YOLOv8 algorithm is used in this research and by using transfer learning, its pre-trained weights act as a backbone for training the custom dataset. Like other versions of YOLO, it detects objects by using six important attributes coupled with bounding box coordinates values. Important attributes of an object detection API include a bounding box such as the top left corner of the x and y-coordinates, its width and height, a confidence score that has a probability of 0 to 1, or a class ID or tag. This work contributed to higher detection accuracy in office surveillance systems using YOLOv8 which was an ideal algorithm for detecting objects from a performance-based point of view.

Training on custom data and all related operations were supported by the YOLOv8 GitHub repository by storing the training weights of the model and accuracy, etc. Transfer learning can be conducted using six different variants of pre-trained weights while custom dataset training. The training weights for object detection included "yolov8n.pt" for nanoscale objects, "yolov8s.pt" for mini objects, etc., up to, "yolov8x.pt" for very huge objects. Selecting the proper weight for final training was also based on accuracy coming from training on these different training weights. In this research, "yolov8n.pt" was used for object detection, and "yolov8n-seg.pt" for Image Segmentation was used for final training to get accuracy by using these base weights. The Furniture dataset is divided into three parts training, validation, and testing. The Furniture dataset is split into training, validating, and testing parts. There are 3617, 1125, and 213 images in each respective part. The image size of the dataset used in this work by Yolov8 is 640 × 640. However, the augmentation process cannot be performed by the YOLOv8 model because it does not have this function. Therefore, we used the Roboflow website which helped us in the processing of the YOLOv8 method. Roboflow's website helped us with data preprocessing and augmentation techniques, including resizing, horizontal and vertical flips, rotations, noise, and more. In the Person dataset, there are 877 images in training data, 194 in validation data, and 213 in testing data. After the preprocessing stage, the Yolov8 model is applied to the Training set.

The Furniture dataset was trained for 100 epochs while using "yolov8n.pt" as the training weights for object detection. The Person dataset was trained for 16 epochs while using "yolov8n.pt" as the training weights for object detection. The Furniture dataset is split into training, validating, and testing parts. For image segmentation, only 1454 images out of the 3617 are used for training and testing. After the preprocessing stage, the Yolov8 model is applied to the Training set. For Image Segmentation, some selective images from the Furniture Dataset were resized to 640 × 640 pixels and training was done with 16 epochs while using "yolov8n-seg.pt" as the training weights for Image Segmentation. After the completion of the training, the corresponding validation data was used to check the performance of the training phase.

The performance understanding of these data versions was thus facilitated by these results. Thereafter, the saving of train weights was done and the outcomes were measured on test datasets. Since it was this, the weights were done onto the validation data corresponding to the training and after testing the related test set. Subsequently, the saved weights were also tested by using the validation and test datasets. The proposed Methodology using YOLOv8 is shown in Figure 2, in which firstly both the custom datasets are split into training, validating, and testing sets. Then the YOLOv8 model is applied to the training set. After training, the model will come up with custom weights. These custom weights are applied to the validation and testing sets for the evaluation of the model. After the Evaluation step, the results are undergoing through the feature extraction model which is used to extract the coordinate values of the detected objects. In the next step, the cropping function is applied to the Person Dataset to crop the area of interest that can be used in the future to predict the face of the person in case of any anomaly. On the other hand, the overlapping algorithm is applied to the extracted feature to check the overlapping between both the custom datasets. If any change occurs in the office tables then we can track the detected person because we already cropped the area of interest of the Person Dataset.

4. Results and Discussion

In this work, overall implementations were carried out under Python's umbrella as part of Google Colab's scope. Furthermore, as far as transfer learning is concerned, the YOLOv8 algorithm-enabled training is based on the custom dataset and its train weights. We used the Furniture and Person datasets as custom datasets. The stored training weights of each of these custom datasets were applied in evaluation by different versions of the validation and testing datasets. This study was evaluated using mAP (mean Average Precision) as a metric, mAP50, and mAP50-95 are its two predefined outcomes, implemented by the YOLOv8 model. To calculate the detected bounding box overlap between the ground truth boxes and predicted bounding boxes, mAP is used as the intersected over union in this case. Considers the true positive detection of the corresponding detected bounding boxes where the IoU value is more than 0.5. MAP50-95 is also considered as a different threshold ranging between 0.5 and 0.95.

The typical mode of communicating results was mPA50, while mAP50-95 was seldom applied unless in an expressly mentioned section. These sequences were undertaken to select the best weights for training. YOLOv8 algorithm has no image augmentation techniques, which is left to a third party like Roboflow. We tried different augmented variations using Roboflow and selected the best among them by examining its validation result in training. Finally, we separately estimated the training efficiency of the Furniture and Person datasets. The YOLOv8 train process works on any kind of image version regardless of their size while changing all custom datasets images into 640×640 pixels and then giving these images to a training

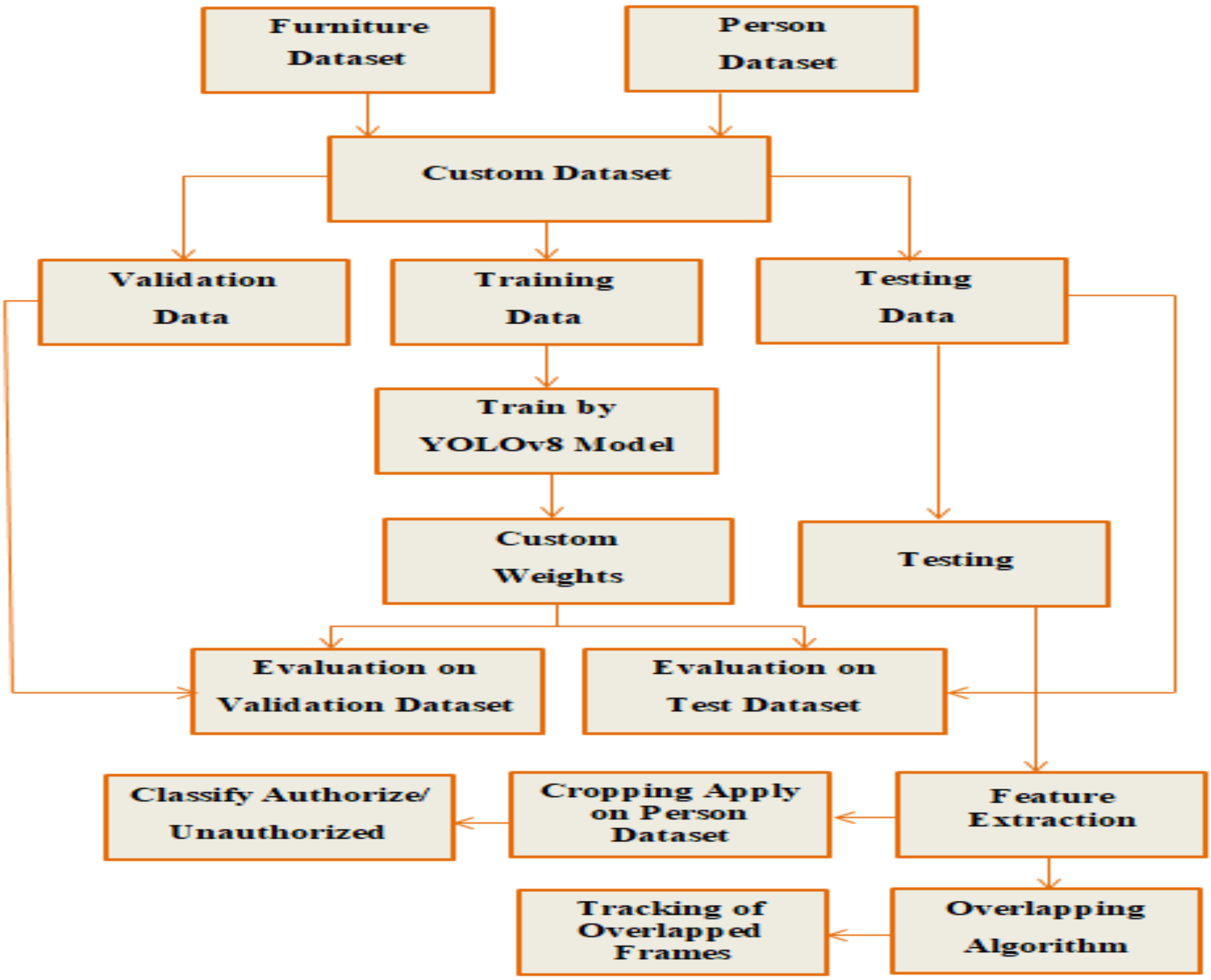
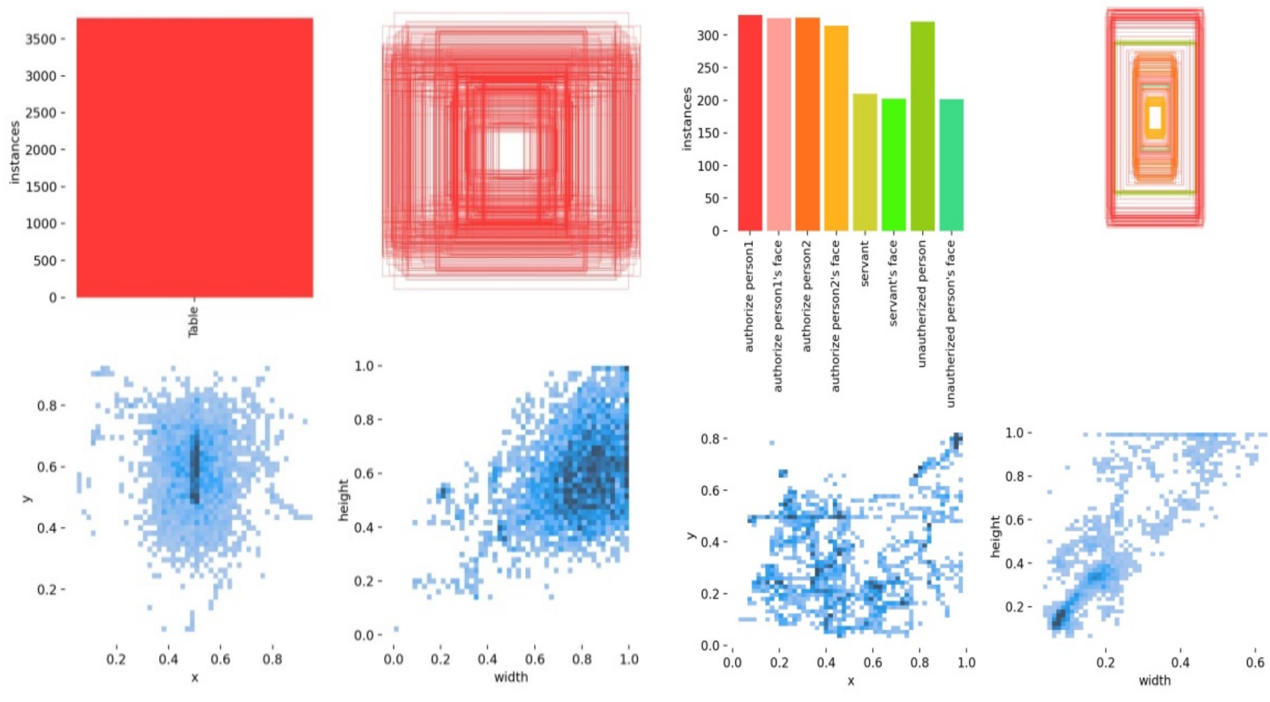


Figure 2. Proposed Methodology using YOLOv8

4.1 Using best Weights for the evaluation of the Furniture and Person Datasets

During the training phase, best weights are evaluated from the YOLO8 model to check the evaluated accuracy of detected objects on validating and testing datasets of the Table and Person images. Here the detection performance outcome for the Furniture and Person datasets are shown by graphs in Figure 3 (a). In the below section, Figure 3 (a) presents the performance of the Furniture dataset while Figure 3 (b) shows the performance of the Person dataset. Also, YOLOv8 uses the best weights to check the performance of

Image Segmentation on validating and testing datasets of the Table images. The segmentation performance outcome for the Furniture dataset is shown in the graph in Figure 4.



(a) (b)

Figure 3. Detection performance using YOLOv8 on custom Dataset (a) Using the Furniture Dataset (b) Using the Person Dataset

Precision-recall (PR) curves are presented in Figure 5 that used the best training weights that were trained on all the images of the Furniture and Person datasets individually. Figure 5 (a) shows the mAP50 score of 99.3% when training weights of the Furniture dataset are applied to the testing set. Similarly, Figure 5 (b) shows the mAP50 score of 70.1% on all classes when training weights of the Person dataset are applied to the testing set. The Mask Precision-Recall (Mask PR) curve and Box Precision-Recall (Box PR) curve shown in Figure 5 (c, d) showing the mAP50 score of 97.4% and 98.6% respectively when training weights of the Furniture dataset are applied to the testing set for segmentation purposes.

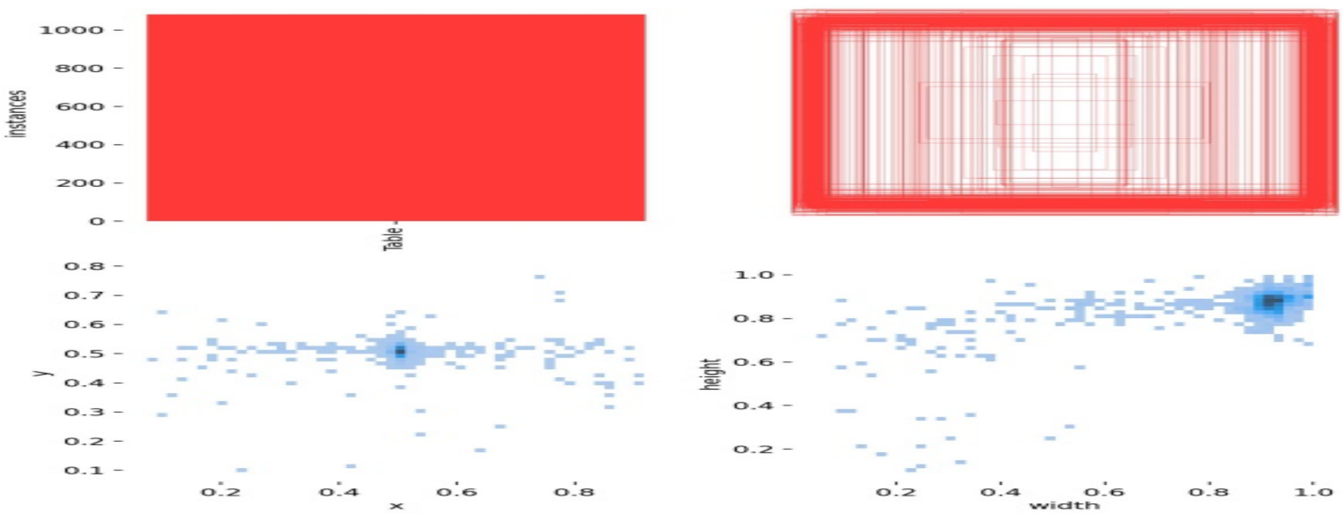


Figure 4. Segmentation performance using the Furniture Dataset

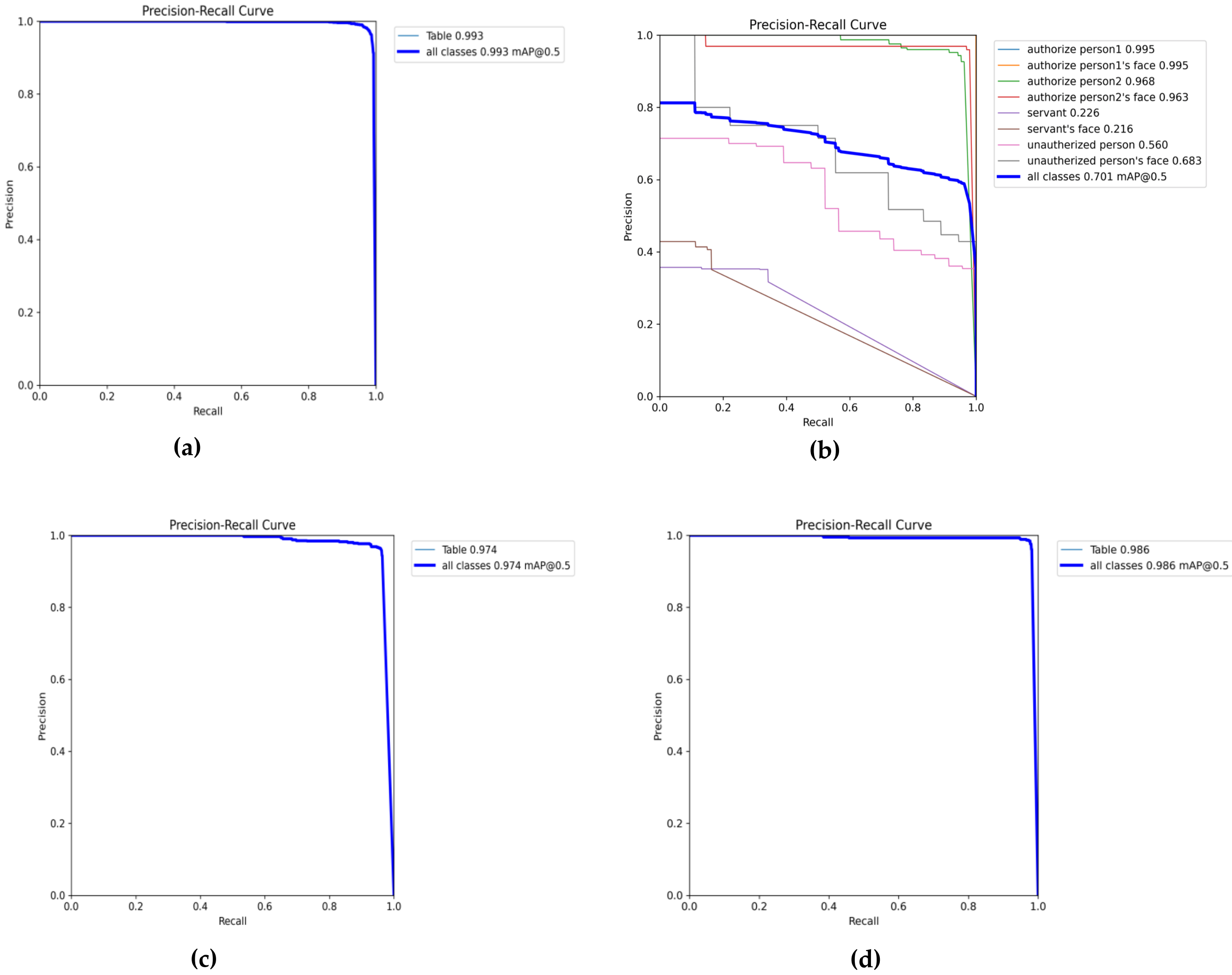


Figure 5. Precision-recall (PR) rate using several training weights (a) (PR) of the Furniture dataset results in detection (b) (PR) of the Person Dataset results in detection (c) (Mask_PR) of the Furniture dataset results in segmentation (d) (box_PR) of the Furniture dataset results in segmentation.

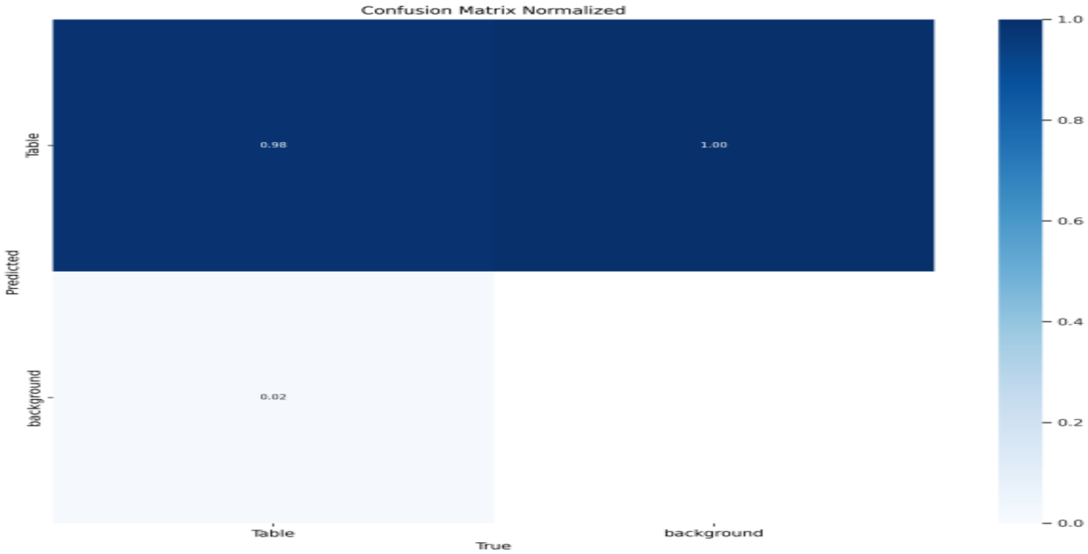
Precision is the total number of true predictions. While the model's sensitivity is measured by the Recall. Precision and recall can be calculated by following the formulation:

Precision = TP/(TP + FP)

While

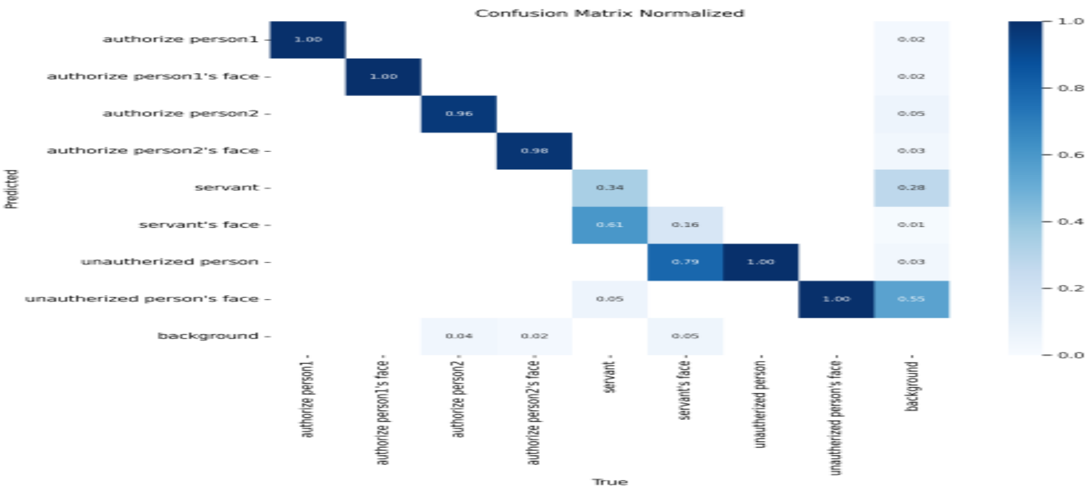
Recall = TP/(TP + FN)

The normalized confusion matrix graphs show the distribution of the predicted class labels in Figure 6 relative to the true class labels in the detection and segmentation problem. The true positives, true negatives, false positives, and false negatives values of the labeled classes are shown pictorially using the training weights of the various datasets. Figure 6 (a) shows the normalized confusion matrix graph for the Furniture dataset to detect tables, Figure 6 (b) shows the normalized confusion matrix graph for the Person dataset to detect persons, and Figure 6 (c) shows the normalized confusion matrix graph for the Furniture dataset to segment tables.



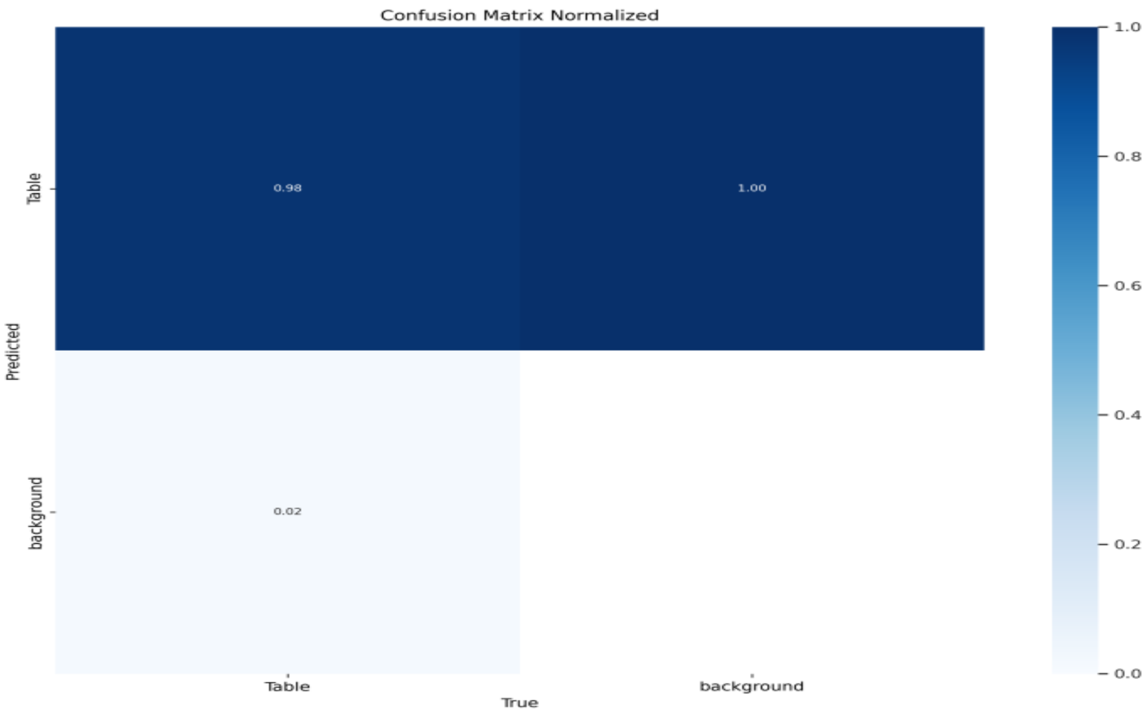
Confusion Matrix Normalized
Table
background
Predicted
0.98
1.00
0.02
Table
background
True
1.0
0.8
0.6
0.4
0.2
0.0

(a)



Confusion Matrix Normalized
authorize person1
authorize person1's face
authorize person2
authorize person2's face
servant
servant's face
unautherized person
unautherized person's face
background
Predicted
True

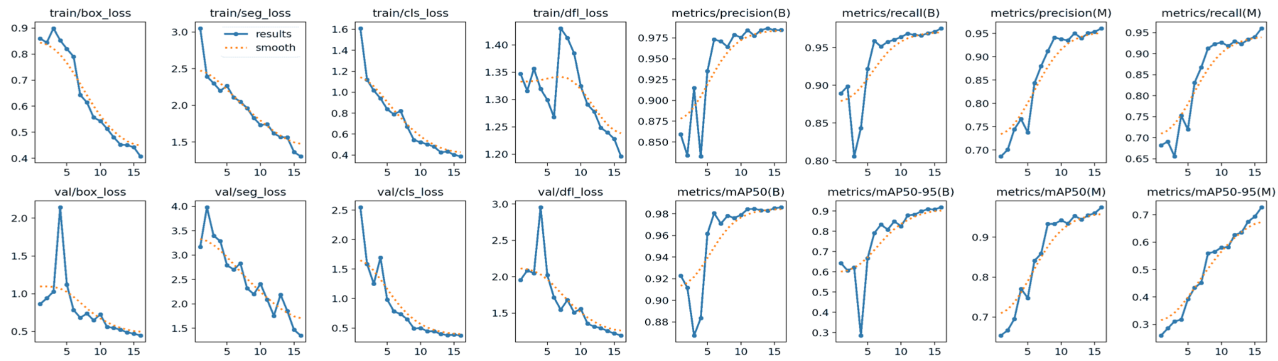
(b)



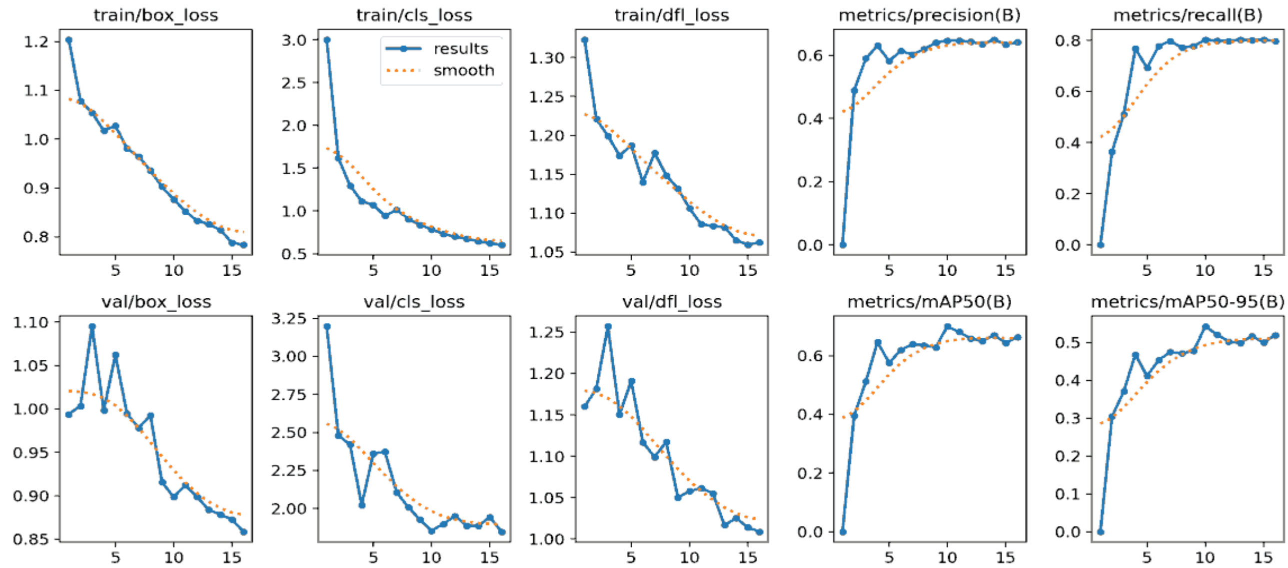
Confusion Matrix Normalized
Table
background
Predicted
0.98
1.00
0.02
Table
background
True
1.0
0.8
0.6
0.4
0.2
0.0

(c)

Figure 6. Normalized confusion matrix graphs (a) To Furniture dataset as a result of detection (b) To Person dataset as a result of detection (c) To Furniture dataset as a result of segmentation



(a)



(b)

Figure 7 (a, b) is used to show the results for detection on the Furniture and Person dataset, and Figure 7 (c) is used to present the segmentation result on the Furniture dataset. The YOLOv8 model shows better results than the results discussed in the recent studies [28]. Here is the graphical representation of the results that shows the curves of the box_loss, cls_loss, dfl_loss, precision (B), and recall (B) and many more.

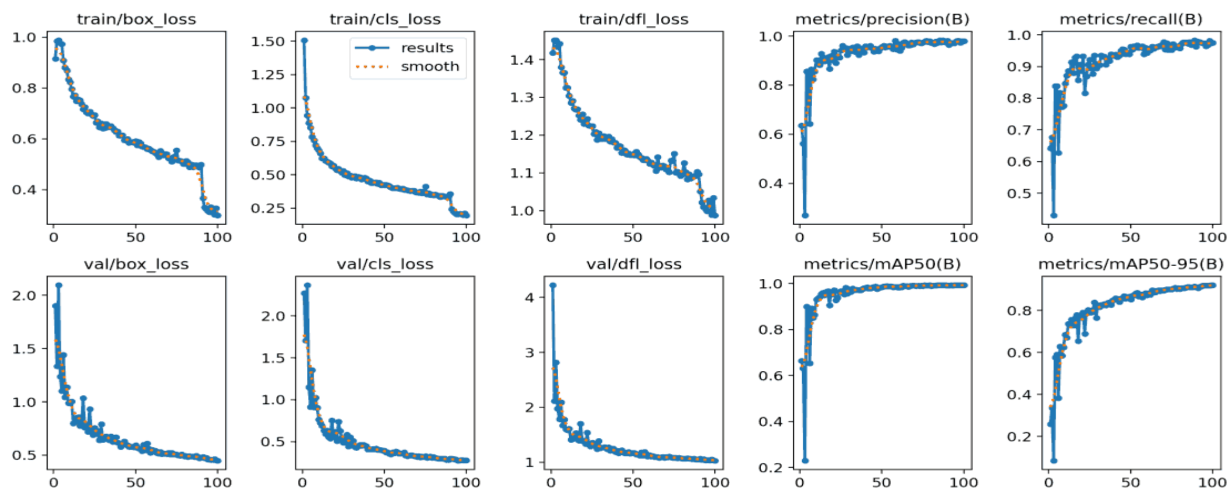


Figure 7. Results of best training weights of the various custom datasets.

Object detection and Image Segmentation reports are discussed in this session, by using the YOLOv8 model. Table 1 and Table 2 present the detection report on the Furniture and Person Dataset respectively while Table 3 presents the Image Segmentation report on the Furniture Dataset.

Table 1. Detection Report for the Furniture Dataset

Class	Images	Instances	Box(P)	Recall	mAP50	mAP50-95
all	1125	1164	97.6%	97.9%	99.3%	92.1%

Table 2. Detection Report for the Person Dataset

Class	Images	Instances	Box(P)	Recall	mAP50	mAP50-90
All	215	557	0.647%	0.801%	0.701%	0.543%
authorize person1	215	100	0.998%	1	0.995%	0.823%
authorize person1's face	215	96	0.989%	1	0.995%	0.795%
authorize person2	215	105	0.943%	0.947%	0.968%	0.688%
authorize person2' face	215	97	0.969%	0.959%	0.963%	0.653%
Servant	215	38	0.329%	0.342%	0.226%	0.188%
Servant's face	215	80	0.4%	0.163%	0.216%	0.15%
Unauthorized person	215	23	0.284%	1	0.56%	0.54%
Unauthorized person's face	215	18	0.264%	1	0.683%	0.51%

Table 3. Segmentation Report for the Furniture Dataset

Class	Images	Box(P)	Recall	mAP50	mAP50-95	Mask(P)	R	mAP50
All	437	0.984%	0.975%	0.986%	0.918%	0.961%	0.96%	0.974%

Table 4, Table 5, and Table 6 represent all the speeds like preprocessing, inference, and post-process per image is mentioned about detection and segmentation on the Furniture and Table Dataset.

Table 4. Speed and inference Report for the Furniture Dataset (Detection)

Preprocess Speed (ms)	Inference (ms)	Postprocess per Image (ms)
0.2	2.5	3.0

Table 5. Speed and inference Report for the Person Dataset (Detection)

Preprocess Speed (ms)	Inference (ms)	Postprocess per Image (ms)
3.2	3.3	4.4

Table 6. Speed and inference Report for the Furniture Dataset (Segmentation)

Preprocess Speed (ms)	Inference (ms)	Postprocess per Image (ms)
1.4	3.3	1.5

4.2 Feature Extraction and Cropping

Extracting features from detected objects using the YOLOv8 model is crucial for predicting new images as shown in Figure 8. By extracting all the coordinate values of the detected objects from the Furniture and Person dataset, we can make data-driven decisions based on the change in object shapes over time. A log file is created to save the extracted object values, which can be useful for future analysis. To streamline the process, the cropping function of the YOLOv8 model is applied to the Person dataset. This function crops all the detected images and saves them into authorized, unauthorized, and servant folders individually, along with their detected faces.

Figure 8. Detected and Segmented annotated frames on custom dataset.

Figure 9. Detect person is cropped from the Person Dataset to check the correlation between consecutive frames

This feature helps to categorize and organize the images, allowing for easy identification and analysis of specific object shapes. Cropping results from the Person Dataset to check the correlation between consecutive frames to secure the office place are shown in Figure 9.

Figure 10. Detect person is cropped from the Person Dataset to check the correlation between consecutive frames

4.3 Overlapping between both the datasets

It is necessary to think about how existing data interact with each other when working with multidirectional datasets and why this interaction causes a bias while forming a generalized statement. In the long run, this could compromise the performance of deep learning models as shown in **Figure 10**. On that note, it is important firstly to scrutinize the data points of intersection and secondly to establish the common values in incongruent sets of data. For this reason, we employ an incremental feature extraction approach which keeps track of every value in a log file. Subsequently, we utilize an overlapping method to illustrate locations where the values for both sets fall on one another, represented graphically for straightforward understanding.

It involves various preprocessing steps such as turning log files into comma-separated values for uniformity purposes between both datasets. Thoroughly analyzing similar data sets may reduce bias and enhance the deep learning model's output. The Collision algorithm is as follows to find the overlapping between the detected frames: Figure 10 shows the overlapping between the table and annotated frames of the Person dataset pictorially. Table frames are shown by green rectangles while overlapping with person frames can be detected by red rectangles. Then we can monitor the activities of each person and store them in the database using the log file. We will be able to match the changes in the log file against the images stored in the database folder for the identification of different persons behind the activities.

5. Conclusions

This research is innovative in improving object boundary identification and feature extraction application through YOLOv8. We were able to obtain very good results by concentrating on the Furniture Dataset including precision for bounding box prediction, mAP50, mAP50-95, and recall with an accuracy of 97.6%, 99.3%, 92.1%, and 97.9%, respectively, which are used to detect and monitor tables under varied office conditions. Lastly, we demonstrated improved performance of the YOLOv8 model to the Person Dataset for the detection of different human activities without infringing on a person's confidentiality. The creation of a database to record every change in any object including the movement or non-movement of persons is very crucial to office security. Feature extraction not only helps real-time decision-making, but it is also useful as input data for future research. The cropping feature helps ease the arrangement of the images thereby enabling one to locate distinctive object shapes within the frame, especially about actions involved in humans. We deal with the issue of formulating generalized sentences by looking at converging sets of data points of intersections and shared values. Visual representation of locations where datasets overlap gives a clear picture of data relations leading to fairness in deep learning model operation. Office security however involves more than just object detection and overlapping algorithms. It is important to investigate other reliable ways of detecting overlaps in datasets in a bid to achieve real safety in the workplace. It helps us be just a step ahead of any possible security menace to ensure that an office environment is secure for all.

References

1. Xie, Q., et al., Object detection and tracking under occlusion for object-level RGB-D video segmentation. IEEE Transactions on Multimedia, 2017. 20(3): p. 580-592.
2. 2. Felzenszwalb, P.F., et al., Object detection with discriminatively trained part-based models. IEEE transactions on pattern analysis and machine intelligence, 2009. 32(9): p. 1627-1645.
3. 3. Uijlings, J.R., et al., Selective search for object recognition. International journal of computer vision, 2013. 104: p. 154-171.
4. Deng, C., et al., Extended feature pyramid network for small object detection. IEEE Transactions on Multimedia, 2021. 24: p. 1968-1979.
5. Qiao, S., L.-C. Chen, and A. Yuille. Detectors: Detecting objects with recursive feature pyramid and switchable atrous convolution. in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021.
6. Wang, W., et al. Internimage: Exploring large-scale vision foundation models with deformable convolutions. in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023.
7. Zhang, H., et al., Dino: Detr with improved denoising anchor boxes for end-to-end object detection. arXiv preprint arXiv:2203.03605, 2022.
8. Taye, M.M., Understanding of Machine Learning with Deep Learning: Architectures, Workflow, Applications and Future Directions. Computers, 2023. 12(5): p. 91.
9. Kamath, V. and A. Renuka, Deep Learning Based Object Detection for Resource Constrained Devices-Systematic Review, Future Trends and Challenges Ahead. Neurocomputing, 2023.
10. González, J.L.S., et al., Real-time gun detection in CCTV: An open problem. Neural networks, 2020. 132: p. 297-308.
11. Kang, J., et al., A survey of deep learning-based object detection methods and datasets for overhead imagery. IEEE Access, 2022. 10: p. 20118-20134.
12. Redmon, J., et al. You only look once: Unified, real-time object detection. in Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.
13. Redmon, J. and A. Farhadi. YOLO9000: better, faster, stronger. in Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.
14. Redmon, J. and A. Farhadi, Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767, 2018.
15. Liu, W., et al. Ssd: Single shot multibox detector. in Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14. 2016. Springer.
16. Lin, T.-Y., et al. Feature pyramid networks for object detection. in Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.
17. Lim, J., et al. Gun detection in surveillance videos using deep neural networks. in 2019 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC). 2019. IEEE.
18. Ingle, P.Y. and Y.-G. Kim, Real-Time Abnormal Object Detection for Video Surveillance in Smart Cities. Sensors, 2022. 22(10): p. 3862.
19. Ahn, H. and H.-J. Cho, Research of multi-object detection and tracking using machine learning based on knowledge for video surveillance system. Personal and Ubiquitous Computing, 2022: p. 1-10.
20. Osipov, A., et al., Deep learning method for recognition and classification of images from video recorders in difficult weather conditions. Sustainability, 2022. 14(4): p. 2420.
21. Rani, S., D. Ghai, and S. Kumar, Object detection and recognition using contour based edge detection and fast R-CNN. Multimedia Tools and Applications, 2022. 81(29): p. 42183-42207.
22. Ekanayake, B., et al., A deep learning-based approach to facilitate the as-built state recognition of indoor construction works. Construction Innovation, 2022.
23. Sharma, T., et al., Deep learning-based object detection and scene perception under bad weather conditions. Electronics, 2022. 11(4): p. 563.

24. Wang, G., H. Zheng, and X. Zhang, A robust checkerboard corner detection method for camera calibration based on improved YOLOX. Frontiers in Physics, 2022. 9: p. 828.
25. Berardini, D., et al., A deep-learning framework running on edge devices for handgun and knife detection from indoor video-surveillance cameras. Multimedia Tools and Applications, 2023: p. 1-19.
26. Gautam, A., et al., Deep learning approach to automatically recognise license number plates. Multimedia Tools and Applications, 2023: p. 1-18.
27. Deepa, D. and A. Sivasangari, An effective detection and classification of road damages using hybrid deep learning framework. Multimedia Tools and Applications, 2023. 82(12): p. 18151-18184.
28. Kumar, D. and N. Muhammad, Object detection in adverse weather for autonomous driving through data merging and YOLOv8. Sensors, 2023. 23(20): p. 8471.
29. Deng, J., et al. Imagenet: A large-scale hierarchical image database. in 2009 IEEE conference on computer vision and pattern recognition. 2009. Ieee.
30. Chen, D., et al., Anomaly detection in surveillance video based on bidirectional prediction. Image and Vision Computing, 2020. 98: p. 103915.
31. Wang, S., et al., A moving track data-based method for gathering behavior prediction at early stage. Applied Intelligence, 2021. 51(11): p. 8498-8518.
32. Tao, W., et al., Worker activity recognition in smart manufacturing using IMU and sEMG signals with convolutional neural networks. Procedia Manufacturing, 2018. 26: p. 1159-1166.
33. Patrikar, D.R. and M.R. Parate, Anomaly detection using edge computing in video surveillance system. International Journal of Multimedia Information Retrieval, 2022. 11(2): p. 85-110.
34. Yadav, P., N. Gupta, and P.K. Sharma, A comprehensive study towards high-level approaches for weapon detection using classical machine learning and deep learning methods. Expert Systems with Applications, 2023. 212: p. 118698.