OccupEye -Building Traffic and Animal Monitoring System

T Gopi Manohar Reddy[1], M Sai Sanjana Reddy[2], K Koumudi Reddy[3], K Rupesh Sai Narendra[4] , K. Srinivasa Reddy[5]

[1,2,3,4,5]School of Computer Science & Engineering, VIT-AP University, Amaravati, Andhra Pradesh, India.

gopimanohar.20bci7087@vitap.ac.in , sanjana.20bci7057@vitap.ac.in , koumudi.20bci7180@vitap.ac.in
rupeshsai.20bcn7089@vitap.ac.in , srinivasareddy.k@vitap.ac.in

Abstract. In our rapidly evolving world, precise people counting and detection within buildings have become more important than ever. With a global population of over 7.7 billion people, cities have become busy centers of activity. At the same time, smart cameras and surveillance systems have ushered in a new era. Counting people isn't just about attendance tracking, it's also about optimizing resources and improving safety protocols. For example, shopping malls can adapt trading hours based on actual foot traffic. The convergence of people-counting technologies and animal detection technologies is changing the way we interact with our environment, making it smarter, safer, and more efficient. Thanks to computer vision and AI, these innovations are highly accurate, cost-effective, and non-intrusive. In addition, the detection of animals entering buildings is essential, especially as interactions between humans and animals become more commonplace. Buildings that detect animals such as rodents or unwanted wildlife will trigger appropriate responses to protect human health and property. These technologies provide valuable insights to businesses, educational institutions, and urban planners, allowing them to make informed decisions and improve safety measures. They are not just technological marvels, but transformational forces that redefine our relationship with our spaces and enrich our daily lives.

Keywords: Computer Vision, People Counting, Building Surveillance, Animal Detection

1 Introduction

The Evolution of the world is more advanced in Technology as well as in Population. Control of the growth of the population is very hard, but we use technology to control the whole population to gather in one place. We humans have a psychological factor to celebrate or spend a good time with loved ones which might be two or three or many more. There will be a major problem if the number of individuals increases at the same place. The consequences of the power of the crowd will lead to destruction [1] like a family of 15 people being at the same place and another family with the same strength coming to meet each other, this may lead to major safety factors like building Infrastructure, ventilations, building capability, fire exits, etc.

Large gatherings always have a vital role in the life of people, but that is not at present movement. One of the key characteristics of the modern era is the replacement of individual

ICISML 2024, January 05-06, Pune, India
Copyright © 2024 EAI
DOI 10.4108/eai.5-1-2024.2342545

conscious activity with the unconscious action of crowds [1]. Crowd catastrophes, in which people suffer severe injuries or lose lives as a result of being crushed or tramped on, are not only caused by emergencies like fires, violent crowd conditions, or the excessive elation of some crowd members. Such incidents can happen everywhere, including sporting events, religious ceremonies, and rock concerts. [2]. Disasters caused by crowding have been defined by Fruin as a specific type of pedestrian traffic process in which certain critical performance limits have been exceeded. Fruin defines crowding as the sudden gathering of a large number of people in an enclosed space with sufficient mass and force to cause human injury or death even with the latest common problem of Covid pandemic[3][4].

Proper arrangements should be arranged or backups have to be kept to control the catastrophe cases of the crowd disasters[12]. Here are some of the records, on 11 October 1711, a collision between a carriage and a cart led to trap a large crowd in the middle of the Guillotière bridge in Lyon, France, which led to the death of 245 people. On May 30, 1770, a firework display commemorating the union of the future King Louis XVI and his consort, Marie Antoinette, in the Place de la Concord in Paris, France, led to the death of at least 133 people. The fire was caused by a malfunctioning mannequin and other decorations, leading to a panic in which many onlookers were crushed beneath their feet and some were drowned in the nearby river. The 1823 Carnival tragedy in Malta, during the celebration of Carnival, resulted in the death of 110 young boys attempting to leave the Brooklyn Theatre after attending a concert featuring the musical group Minori Oservanti. As a result of the tragedy, 278 people lost their lives on December 5, 1876.

Brooklyn Theatre fire occurred on 18 May 1896 where 1389 people died during the coronation of Tsar Nicholas II because of crowd crush. 71 died on 4 March 2010 in Pratapgarh stampede (India), The Phnom Penh (Cambodia) stampede killed 347 people on November 22, 2010, 102 died on 15 January 2011 in 2011 Sabarimala crowd crush (India), On February 1st, 2012, there was a huge disturbance at the Port Said Stadium in Egypt, resulting in the death of 74 people, 242 people lost their lives in the Kiss nightclub fire in Brazil on January 27, 2013. Another 26 people were injured in a stampede that occurred shortly after Dussehra at Gandhi Maidan in Patna, India. 135 died in the Kanjuruhan Stadium disaster, Indonesia on 1 October 2022. Halloween revelers suffered injuries in a small lane in the Itaewon neighborhood; at least 172 additional people were hospitalized. Officials determined that one survivor's suicide death in December 2022 was due to the tragedy, making him the 159th victim by the law in South Korea.

In this study, we will be applying convolution neural network models to see the occupancy of the building from time to time and able to monitor the public flow into and out of the building at the same time, we use the latest version of the CNN(Convolution Neural Network) model YOLO (You Only Look Once) algorithm and the optimizers of our choice SGD, Adam, Adamax, AdamW, NAdam, RAdam, RMSProp based on the best performance to detect the Objects in the frame like persons, dogs, cats, handbags, chairs etc[5]. The continuation of these sections will be Section 2 which addresses the related the related work of our study, Section 3 Methodology shows and explains our approaches and methods of working, Section 4 describes the results and Section Conclusion then we conclude the paper with proper references at the ending.

2 Related Work

In the fields of AI, Deep learning, Computer vision, etc. Object detection is a basic study area where it has a majority of the complex tasks are done by it. It locates the targets based on the area of interest taken by the picture or frames in the video and replies with the bounding boxes[6]. The idea of object detection was first introduced in 2014 with the first pre-sized model, the R-CNN, which had an average presession percentage of objects at the time of capture (mAP) that was 53.7% according to PASCAL's VOC 2010 [7]. The models consist of several parts such as modal design, test-time detection, training, and evaluation, by then we will be able to get the map values based on the evaluated score. In the list of multiple detection modes R-CNN(Region-Based Convolutional Neural Network) , Fast R-CNN, Faster R-CNN, SPP-net, R-FCN, FPN(Feature Pyramid Network) and Mask R-CNN, these are based on region proposal, where some of other model based on classification or regression like DSOD [11], MultiBox, AttentionNet, SSD, G-CNN, YOLOv2 [8], DSSD[10] and YOLO(You Only Look Once) [9].

Among them, in this paper, we are using the regression-based model named YOLO(You Only Look Once) to detect the object and we explore the latest versions of it. This is a regression game in which we get bounding box and class probabilities straight from the image pixel. We use our models to predict the objects from those image pixels. The main reason for choosing this YOLO model is, that this is an extremely fast algorithm as this algorithm detects a regression problem we don't need any complex pipelines when compared with others. YOLO(You Only Look Once) sees the whole frame or picture at the time of training or testing so it interprets the relevant information about the classes and their appearance. When compared to Fast R-CNN(Fast Region-Based Convolutional Neural Network) which is a top detection model, YOLO creates less than half the amount of background mistakes. YOLO can recognize and recognize objects in a wide range of ways. It's way better than other detection algorithms, like DPM or R-CNN(Region-Based Convolutional Neural Network) when it's trained on real photos and compared to artwork. Plus, since it's so versatile, it's less likely to mess up when you're trying to figure out what something looks like or when you enter something unusual[9].

The YOLO framework is composed of five bounding boxes, each of which is expected to have a lattice unit at its focal point. The bounding boxes are expected to be affected by the quality of the entire image. The structure of YOLO is composed of three main components: the backbone, the neck, and the head. For the Aquatic dataset, the YOLOov5 model achieved a mAP of 0.84, the Sign Language dataset of 0.87, the Chessboard dataset of 0.9, and the Library Books dataset of 0.86. This demonstrates that version 5 of YOLO works well for the detection. For the Racoon dataset, this model achieved the highest mAP with an accuracy of 91%. In this paper, we are going to use the latest version of the YOLO(You Only Look Once) which is a YOLOv8 that was launched in 2023 2 years after launching v5, with major improvements and can do detection, segmentation, and classification [13].

In this paper we are going to use YOLO(You Only Look Once) algorithm by considering its speed and accuracy, we are going to perform object detection on each individual frames in the video. As the current running world is using high definition cameras with high frames per seconds such as 30fps(frames per second), 60 fps etc. As we are mainly considering live footage of cctv surveillance camera, we have to take suitable model for processing those multiple frames

without missing any. YOLO performs quicker than other model we known. So, we are going to use YOLO(You Only Look Once) to get fastest results of each frame and able to run work quick and comfortably.

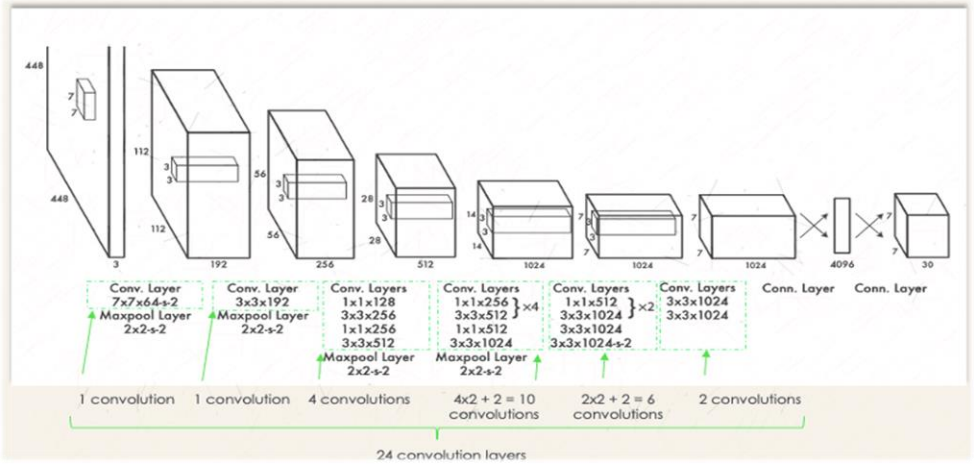

Fig. 1. Architecture of YOLO

3 Methodology

For our operations and performance work we are going to use the YOLOv8 which is the latest version in the YOLO(You Only Look Once) family with cutting-edge performance in accuracy and speed. This latest model was trained on the COCO 2017 Dataset.MS COCO full name Microsoft Common Object in context is a dataset consisting of around 328K Images. When comparing COCO(Common Object in context) with ImageNet, PASCAL VOC 2012, SUN. This additional dataset consists of a variety of datasets, ranging from small to large, with a range of categories and types of images. The primary purpose of PASCAL Visualization and Analysis (PASCAL VOC) is to identify objects in natural images, while SUN focuses on recognizing different types of scenes and the features that are commonly associated with them. MS COCO's purpose is to identify and segment objects as they are encountered in their natural environments[14]. Even though Microsoft Cognos Object-Oriented Computing (COCO) has fewer categories than the SUN and ImageNet, it still has a higher number of predicted instances per category, which will be advantageous for more complex models. Generally, the dataset of COCO typically has a 7.7 instance count per image and a 3.5 instance count per category. On the other hand, ImageNet and the PASCAL Visualization Object Order (POVO) has an average of less than two and three instances respectively per image, as illustrated in Fig.2 [14].

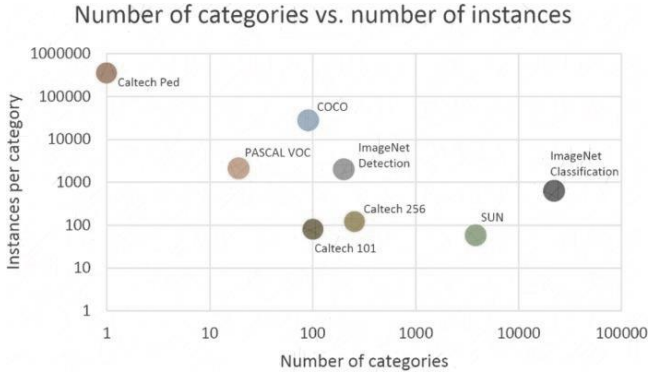

Fig. 2. Number of Category vs Instances on Different Datasets

In this section, we propose that our main work Is on the concept of LOI(Line of Intersection) ,In this method, we are counting the objects that crossed the imaginary line. We track the trajectory of the person using two lines and classify the person as entering or exiting the building. This line of Interest method helps us count the persons, We will be counting the people who pass the imaginary lines fixed at a certain position that everybody must cross through as shown in Fig. 4. The lines are marked using the frame height coordinates and extended to the frame width. Our YOLO(You Only Look Once) model detects the Person or any object in the frame and provides us with the center coordinates of it. We use them to give a unique ID to every object that we are considering. Now these ID plays a key role in our work as we are going to count these IDs as persons and calculate the no of people who entered or exited the building. The Id is determined as Entered or Exited if and only if it crosses both the Imaginary Lines marked by us [16].

Here the geographical angle of the frame plays the key role in differentiating between the entering or exiting, For example, if the camera sensor is installer inside the building, the person walking towards the frame or the ID crossing the Lines toward the frame means they are entering and the Exiting will be the person going away in the frame. And the phenomenon is reversed if the camera sensor is installed outside of the building. Based on this phenomenon we count the people entering or Exiting the building.

In our work other than Persons, animals are also detected and Identified with a unique ID, and now in the case of an animal, if the animal is inside the building or entered the building the Animal is marked with a Bounding box for the whole time it was inside[17]. A System-generated Alert will be called to the corresponding security personnel when any animal enters to building and respective measures will be taken care of [18]. The Measured Data calculated by our algo will be displayed on the screen itself as shown in Fig. 5 in the results section. The same data will be sent to corresponding Officers if the counted persons in the building reach the Limit of the Building capacity based on the architecture of the building. The results and discussion section will discuss all the results with samples.

4 Results and Discussion

The pre-trained model of the YOLO(You only look once) we had used of the latest version 8 named YOLOv8 comes with different variations like nano, small, medium, large, and extra large which was built by the dark web community. This model with the coco dataset got the higher mAP(mean Average Pression) among the other versions with comparing to parameters and Latency A100 TensorRT FP16(ms/img)(Half-precision floating point format – 16 bits) as shown in Fig. 3. [15].

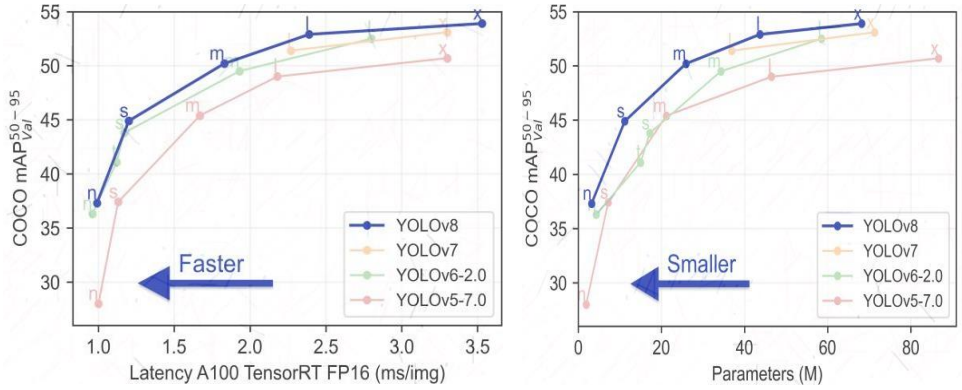

Fig. 3. Representation of mAP Val of different versions of YOLO on COCO

We have used the extra large model for our work as it stands with a high mAP Val among all the variations available for us. As of the Fig.4 we can see the mAP(mean Average Pression) of the Yolov8x model is around 54mAP(mean Average Pression) and the sample result of the model is also shown below.

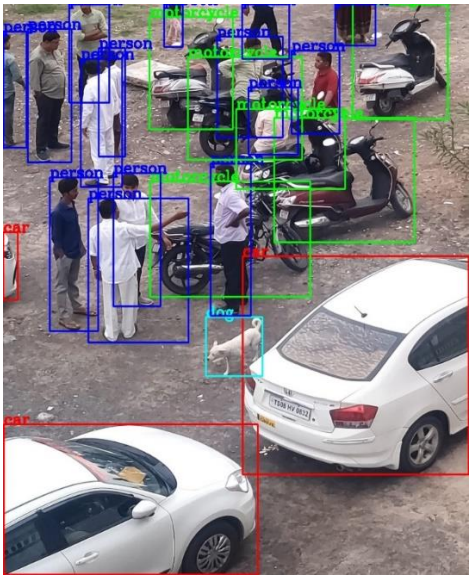

Fig. 4. Performance Output of YoloV8x model on an Image.

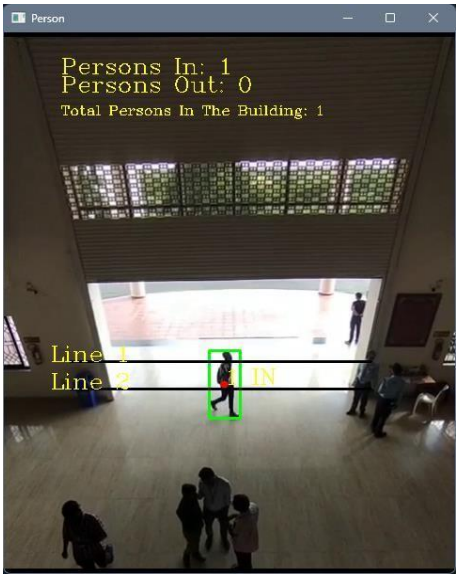

Fig. 5. Entrance of the building and Bounding box around the person while he entered the building.

With all the efforts we finally got the person detected, assigned with unique IDs, and counted when crossed the imaginary lines. The line of intersection phenomenon helps us to determine whether the personnel entered or exited from the building. All the sample images are represented below.

In the Fig 5, we can see that a person is entering a building where there are a lot of people around, Only when the person crosses both lines, then the count be considered, and the Persons inside the building increase by 1. The same process will be continued till the code is stopped manually when it is implemented with the live stream. Here are some more samples.

The Total persons in the Buildings are calculated with the formulae:

$$\textit{Total Persons in The building} = (\textit{Persons In} - \textit{Persons Out}) \quad (1)$$

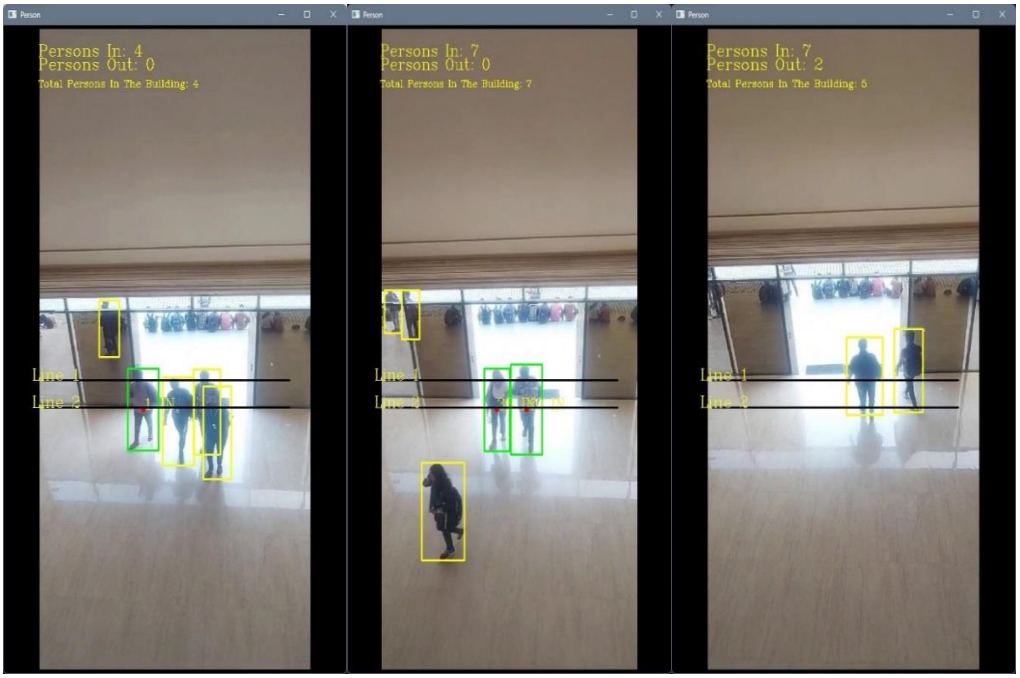

Fig 6. People Entering and Exit the Building.

This images shows that our work here records 7 persons entered the building and 2 persons left the building. So, the total persons sill remain in the building will be 5 which was calculated with the formulae (1). The counting is based on the person crossed the imaginary line that we have set, not based on number of objects or persons detected in the frame.

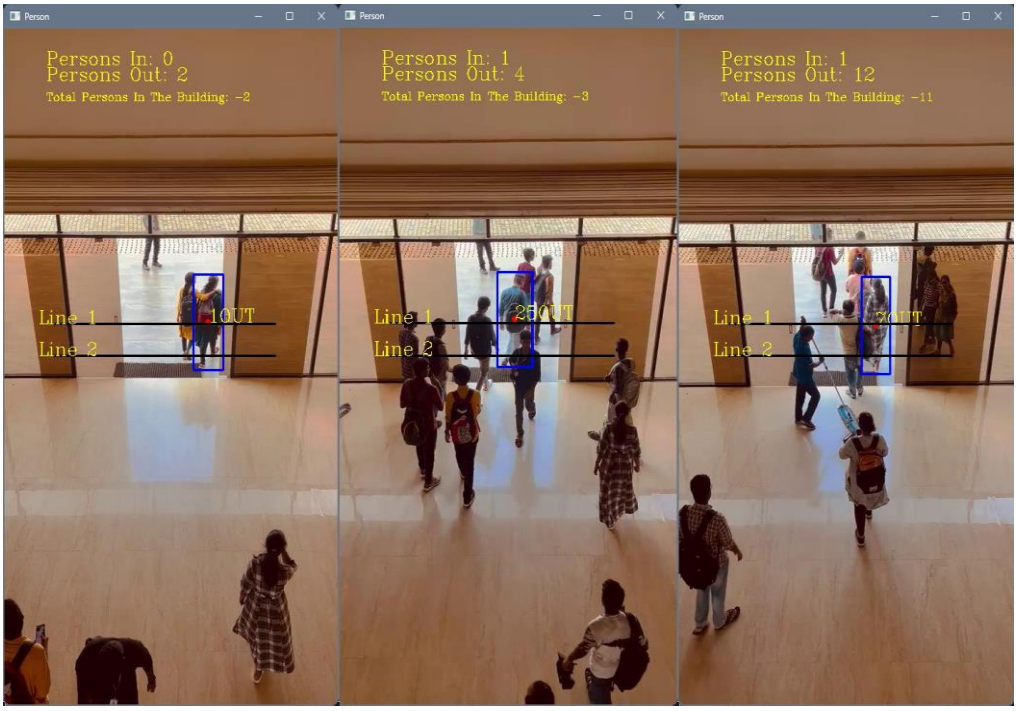

Fig.7: People exiting the building.

People are exiting the building rapidly within seconds, but our model is capable enough to keep an eye on every person on the way. From Fig. 7, 12 personals exited whereas only 1 person entered, the final persons inside the building were in negative which was calculated with the formulae(1) , This type of scenario will be caused when the building has multiple entrances and the video source is from the middle of the rush hour. When the animals entered the building the bounding box was shown in red for the whole time until the animal was sent out. The same is shown in the figure below.

Fig. 8. Animals entering a store to have some food and detected along with a visual representation of a Bounding Box around the animal as long as it stays inside.

In this sample, while the dogs entered the store for food they were marked with a Red Bounding box, which is a visual alert to the Shopkeeper and it is marked with the name of that animal Eg: "Dog IN ALERT". This alert happens when ever an animal's centroid is lower then the lowermost Imaginary line in the frame. And a sound alert will also be triggered to the shopkeeper to alert him about the dogs or animals which will be implemented in the future.

5 Conclusion

By the Year 2023, the world will be advancing with Artificial Intelligence, where every machine helps humans a lot. We managed to perform the detection and surveillance operation using the latest version of the YOLO i.e., YOLOv8. The impact of YOLOv8 extends far beyond crowd management, it has ushered in a new era of computer vision applications. These applications have found their place in a myriad of public spaces, such as stadiums, airports, and event venues, ensuring that the safety and comfort of attendees are prioritized. This remarkable technology has allowed us to efficiently monitor and record accurate crowd data. The versatility of YOLOv8 is further evident in its ability to detect intrusions. This capability has proven invaluable in alerting the presence of animals or unauthorized individuals within buildings and restricted

areas, fortifying security measures and deterring unwanted disruptions. By the work that we have done, people can be saved by restricting them to gather at the same place with huge strength. We managed to bring the computer vision application with the detection algorithm and successes by monitoring the accurate values of the strength of a crowd. This type of application can be very useful in all public places, where a lot of different people join together. Recently we have had so many incidents where animals attacked local villagers. Through this project, We successfully managed to alert the presence of animals inside the building or place, which will be useful in avoiding unwanted disruption and human safety.

In the future, we are going to take this work to the next level by connecting multiple video sources & processing simultaneously, $3D$ localization, and a depth camera will be developed. Also, more advanced DL object detectors will be trained in our custom dataset and compared with YOLOv8. All entrances are tracked and can be processed with a single system and we plan to connect the feed directly to the cloud infrastructure, where the data will be directly reflected on the web portal we are planning to design and all the statistics of the entry flow will be recorded and presented in the form of Graphs and even available to filter for research purpose. The alerts will be called as of two forms, one is as a pop-up on the user interface we are going to create and another with the sound system we will be connect externally to the system, so that all people around will be notifies not only the one how monitors. This Data can be shared with the relevant government departments such as Police, firefighters, etc., where they can also perform manual security monitoring of the building. It's important to underline that while AI, particularly YOLOv8, plays a pivotal role in security and surveillance, it is not a replacement for human oversight. Instead, it complements manual security monitoring, offering real-time insights and alerts to security personnel, resulting in a more robust security infrastructure.

In conclusion, YOLOv8 and AI technologies have not just evolved but revolutionized our surveillance and security systems, significantly advancing our ability to manage crowds, detect intrusions, and maintain public safety. The future holds the promise of even more seamless integration, enhanced data accessibility, and deeper collaboration with government agencies, underscoring the transformative power of AI in enriching lives and ensuring the safety of communities worldwide.

References

1. Borch, Christian. "Crowds, Race, Colonialism: On Resuscitating Classical Crowd Theory." Social Research: An International Quarterly 90.2 (2023): 245-269.
2. Haghani, Milad, and Ruggiero Lovreglio. "Data-based tools can prevent crowd crushes." Science 378.6624 (2022): 1060-1061.
3. Haghani, Milad. "Crowd dynamics research in the era of Covid-19 pandemic: Challenges and opportunities." Safety science 153 (2022): 105818.
4. Khan, Muhammad Asif, Hamid Menouar, and Ridha Hamila. "Revisiting crowd counting: State-of-the-art, trends, and future perspectives." Image and Vision Computing (2022): 104597.
5. Singh, Vikash, et al. "Image/Object Detection Using Shot Multi-Box Detector (SSMBD) Algorithm."

6. Ding, Ning, Ce Zhang, and Azim Eskandarian. "SalienDet: A Saliency-based Feature Enhancement Algorithm for Object Detection for Autonomous Driving." IEEE Transactions on Intelligent Vehicles (2023).Ross Girshick, Jeff Donahue, Trevor Darrell, Jitendra Malik, UC Berkeley: Rich feature hierarchies for accurate object detection and semantic segmentation
7. Sinha, Ankit, Soham Banerjee, and Pratik Chattopadhyay. "Effective Stacking of Deep Neural Models for Automated Object Recognition in Retail Stores." International Journal of Computer and Information Engineering 17.6 (2023): 374-381.
8. Anwar, Masrur, Yosi Kristian, and Endang Setyati. "Klasifikasi Penyakit Tanaman Cabai Rawit Dilengkapi Dengan Segmentasi Citra Daun dan Buah Menggunakan Yolo v7." INTECOMS: Journal of Information Technology and Computer Science 6.1 (2023): 540-548.
9. Abbasi, Ali, et al. "Sensor Fusion Approach for Multiple Human Motion Detection for Indoor Surveillance Use-Case." Sensors 23.8 (2023): 3993.
10. Tyagi, Bhawana, Swati Nigam, and Rajiv Singh. "Person Detection Using YOLOv3." Soft Computing: Theories and Applications: Proceedings of SoCTA 2022. Singapore: Springer Nature Singapore, 2023. 903-912.
11. Jiang, Tingyao, et al. "An improved YOLOv5s algorithm for object detection with an attention mechanism." Electronics 11.16 (2022): 2494.
12. Chen, Yantai, et al. "Building data-driven dynamic capabilities to arrest knowledge hiding: A knowledge management perspective." Journal of Business Research 139 (2022): 1138-1154.
13. Challagundla, Yagnesh, et al. "Screening of Citrus Diseases Using Deep Learning Embedders and Machine Learning Techniques." 2023 3rd International conference on Artificial Intelligence and Signal Processing (AISP). IEEE, 2023.
14. Mao, Xiaofeng, et al. "COCO-O: A Benchmark for Object Detectors under Natural Distribution Shifts." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023.
15. Raj, Deepa. "Recent object detection techniques: a survey." International Journal of Image, Graphics and Signal Processing 13.2 (2022): 47.
16. Argota Sánchez-Vaquerizo, Javier. "Getting real: the challenge of building and validating a large-scale digital twin of Barcelona's traffic with empirical data." ISPRS International Journal of Geo-Information 11.1 (2022): 24.
17. Bao, Jun, and Qiuju Xie. "Artificial intelligence in animal farming: A systematic literature review." Journal of Cleaner Production 331 (2022): 129956.
18. Shingne, Marie Carmen, and Laura A. Reese. "Animals in the city: Wither the human-animal divide." Journal of Urban Affairs 44.2 (2022): 114-136.