Journal of Robotics and Control (JRC)
Volume 6, Issue 4, 2025
ISSN: 2715-5072, DOI: 10.18196/jrc.v6i4.26396

Non-Intrusive Real-Time Tourist Crowd Monitoring for Overtourism Mitigation using YOLOv8-Based Head Detection and Tracking

Kurnia Wijayanti [1], Giva Andriana Mutiara [2*], Bethani Suryawardani [3], Ersy Ervina [4], Guntur Prabawa Kusuma [5]
[1, 2, 3, 4, 5] Department of Applied Science, Universitas Telkom, Bandung, Indonesia
Email: [1] kurniawijayanti@student.telkomuniversity.ac.id, [2] givamz@telkomuniversity.ac.id,
[3] bethanisuryawardani@telkomuniversity.ac.id, [4] ersyervina@telkomuniversity.ac.id, [5] guntur@telkomuniversity.ac.id
*Corresponding Author

***Abstract*—Overtourism has emerged as a critical issue in popular tourist destinations, often leading to environmental strain, reduced visitor satisfaction, and safety concerns. Traditional methods such as ticket counts, or vehicle estimation fail to provide real-time insights or adapt effectively to dynamic outdoor environments. This study proposes a privacy-aware, real-time visitor capacity monitoring system for smart tourism, utilizing YOLOv8-based head detection and Centroid Tracking to ensure accurate, non-intrusive people counting in dense and complex crowd scenarios. Head detection is employed specifically to preserve personal privacy without compromising on detection performance. The system was trained on a custom dataset comprising over 3,000 annotated frames with diverse lighting conditions, occlusion levels, and viewing angles. Deployment at Wana Wisata Kawah Putih, an open-air tourist destination in Indonesia, demonstrated strong performance with 94.2% accuracy, 95.1% precision, and 90.6% recall, while sustaining >60 FPS for real-time execution. The integration of Centroid Tracking enables lightweight, frame-to-frame identity association with minimal computational overhead, making the system suitable for deployment on moderate-performance hardware. Despite its robustness, the system's performance slightly degrades under extreme weather (e.g., fog, direct glare) and rapid lighting transitions, which remain challenges for visual models. Moreover, the current model requires further evaluation for cross-location generalizability. Future research will explore the integration of predictive analytics for visitor flow forecasting, and further optimization of energy efficiency and adaptive detection under environmental uncertainty. This work contributes a scalable, ethical solution for real-time crowd monitoring to support informed, sustainable tourism management.**

Keywords—YOLOv8; Head Counting; Overtourism Mitigation; Centroid Tracker; Real-Time Visitor Monitoring.

I. INTRODUCTION

The rise in tourist numbers at various destinations in recent years has presented new challenges for area managers. The rise in tourist numbers, when not aligned with sustainable development practices, can lead to overtourism. This phenomenon occurs when visitors perceive that a destination has been excessively frequented, resulting in a loss of its original authenticity [1]. Overtourism, as defined by UNWTO, occurs when visitor numbers exceed a destination's capacity, leading to overcrowding, pressure on infrastructure, and rising tensions between tourists and local communities-while also triggering environmental damage and cultural erosion[2]. The implementation of this system is concentrated in the Kawah Putih tourist forest, as illustrated in Fig. 1.

Fig. 1. Kawah Putih forest tourism

The Kawah Putih tourist forest, recognized as a popular destination, has surged in recent years, straining the site's capacity and environment. This forest faces significant challenges associated with overtourism stemming from the influx of uncontrolled visitors [3]. Uncontrolled crowds can lead to various issues, including discomfort, a decline in the quality of the tourism experience, environmental degradation, and heightened risks to tourist safety stemming from overcrowding [4].

During the 2023 Eid holiday week alone, over 48,000 domestic tourists visited in just ten days (a 44% increase from the previous year)[5]. Peak daily visits exceeded 7,300 people [5], far above what the area can sustainably accommodate. A 2022 carrying capacity analysis estimated that the crater's ecosystem and facilities could realistically support only about 255 visitors per day, with effective capacity as low as 98 per day given transport limitations [6]. This uncontrolled tourist influx contributes to environmental degradation, such as soil erosion, damage to endemic flora, and increased unmanaged waste, posing a serious threat to the long-term sustainability of the site[7]. Moreover, overtourism causes overcrowding, reduces visitor satisfaction, and stresses local ecosystems. Waste management is also challenged, as many tourists leave behind litter (especially plastic waste) that degrades the site's natural beauty [8]. These overtourism pressures threaten both the ecosystem integrity and visitor experience at Kawah Putih, underscoring the need for better real-time monitoring and management of visitor numbers. As noted in [7], continuous exposure of

volcanic areas like Kawah Putih to high-intensity tourism accelerates ecological disruption and can lead to irreversible environmental changes. Therefore, these urgent challenges highlight the necessity of implementing a robust system real-time crowd detection system to support effective and responsive tourism management strategies, ensuring the protection of both ecological and cultural values of Kawah Putih.

Accurate crowd monitoring is crucial for mitigating overtourism impacts, but traditional methods are often inadequate. Currently, the traditional approaches employed to quantify visitor numbers at tourist attractions involve tracking entrance ticket sales, estimating the number of vehicles in parking lots, or conducting manual counts by personnel. Nonetheless, these methods exhibit multiple shortcomings, including inaccuracies in real-time calculations, challenges in managing dynamic situations, and constraints in identifying visitors already present in the tourist area. Moreover, sensor-based counting systems, including infrared or RFID sensor [9] necessitate supplementary infrastructure and frequently exhibit inflexibility when utilized in outdoor settings characterized by dynamic conditions.

In recent years, computer vision-based solutions have emerged to automate visitor counting. Early approaches included classical techniques like frame differencing or background subtraction, which struggle with complex outdoor scenes and moving crowds. More robust solutions leverage object detection models. For example, the YOLO ("You Only Look Once") family of detectors (v3 through v7) has been applied to count people in real time by detecting individuals in video feeds. Prior studies showed that a YOLOv3 model could detect people in live video, but without any tracking it risked counting the same person multiple times [10]. Subsequent implementations improved on this by combining detectors with multi-object tracking algorithms (e.g. SORT or IOU trackers) to follow people across frames. Using trackers with YOLOv4 reduced double-counting and enabled basic crowd analytics [11]. However, even state-of-the-art detectors like YOLOv5 have faced challenges in very dense crowds and occluded conditions, sometimes missing individuals or yielding duplicate counts. Furthermore, many vision-based counting systems have overlooked ethical concerns such as visitor privacy, since identifying or recording individuals can raise data protection issues. Alternative crowd counting techniques using density estimation (regression on crowd density maps) avoid tracking each person, but they require extensive training data and often cannot operate in real time [12]. This background highlights the gap for a more accurate, real-time, and privacy-preserving crowd monitoring approach in tourist environments.

To address these challenges, this study proposes a real-time, privacy-preserving visitor monitoring system that leverages the YOLOv8 object detection model alongside a centroid-based tracker to address the shortcomings of prior solutions. This system utilizing computer vision is essential for the automatic and real-time detection and counting of visitors. A rapidly developing approach involves utilizing computer vision technology for the automatic detection and counting of visitors. This approach employs cameras to gather visual information that is subsequently examined by algorithms powered by artificial intelligence [13]. The advancement of deep learning technology has led to the development of methods like You Only Look Once (YOLO), which demonstrate remarkable effectiveness in detecting objects with both speed and accuracy, even in challenging scenarios involving overlapping objects, low lighting, and high density**.**

This system has the potential to supplant traditional methods or restricted ticket sales that demand significant human resources and often lack precision, while offering rapid, precise, accurate, and cost-effective outcomes [14], [15]. Another technology that can be incorporated with YOLO is the head counting method. This approach is among the most widely utilized techniques for assessing crowd numbers. This approach emphasizes identifying the human head as a key indicator of a person, distinguishing it from methods that involve counting the whole body. The application of bounding boxes for head counting, as opposed to body counting, presents benefits regarding data ethics. This method avoids the collection of intricate details concerning an individual's face or identity, thereby enhancing privacy protection. In certain data privacy regulations like GDPR (General Data Protection Regulation), employing monitoring technology that can directly identify individuals is deemed a breach of privacy unless explicit consent is obtained. Consequently, the head counting method is favored as it solely identifies the presence of individuals while maintaining their anonymity.

Utilizing the head counting method allows the YOLOv8 algorithm to enhance its efficiency in tallying individuals in densely populated spaces, all while maintaining a strong commitment to privacy and ethical considerations in monitoring. Moreover, this approach demonstrates greater efficiency in intricate settings, including crowds with overlapping objects, low lighting conditions, or rapid motion [16]. This method enhances the accuracy and real-time capabilities of the crowd detection system while ensuring a greater sense of ethical responsibility in its use within tourist attractions and other public spaces.

YOLOv8 was chosen as the algorithm due to its superior performance regarding speed and detection accuracy compared to earlier versions. YOLOv8 offers improved accuracy and speed over earlier YOLO versions, which is crucial for handling the challenging conditions at Kawah Putih [17]. This study focusses the detection on human heads (rather than full bodies or faces) to maintain privacy – an approach inspired by prior "privacy-preserving" counting methods that count people without uniquely identifying them. Each detected head is assigned a unique ID and tracked across successive frames using a Centroid Tracker, which performs lightweight frame-to-frame association by calculating the distance between object centroids. This simple yet effective approach ensures that a single individual moving through the camera field is only counted once, without the need for heavy computation or deep feature matching. The YOLOv8-and-tracker framework significantly reduces double counting and enhances count accuracy in real time. Compared to YOLOv5-based detectors, YOLOv8 provides greater tolerance to occlusion and small, distant objects, offering more reliable

detection even for partially visible or far-off visitors. The system operates on a moderately powered device and displays the live count of visitors, giving park managers a real-time view of crowd levels without compromising personal privacy.

This model is engineered to identify a range of objects with minimal latency, rendering it suitable for real-time applications [18]. Furthermore, the capacity of YOLOv8 to modify the model according to specific datasets, like those from a distinct tourist environment, enables the system to tailor itself to the unique conditions encountered in the field [19]. The detection outcomes from YOLOv8 will undergo processing through the OpenCV library, enabling the image to be manipulated to showcase the detection results by highlighting each identified head with a box notation. OpenCV was selected due to its capability to handle a range of frame processing and image manipulation that align with the requirements of the system. This system can be utilized to tally visitor numbers, track density levels, and issue early alerts when capacity nears a predetermined limit in the realm of tourist attractions.

Despite these advancements, there are key limitations of deploying an AI-powered counter in the field. Real-time vision models like YOLOv8 require substantial computing power, which can be challenging to provide on energy-efficient or battery-operated hardware in remote sites. Ensuring the system runs on a low-power device (e.g. an NVIDIA Jetson) may necessitate optimizing the model or sacrificing some accuracy for speed. The outdoor environment of Kawah Putih also introduces unpredictable factors – lighting can vary from bright glare to shadow, and the crater is often shrouded in fog or mist that can obscure cameras [20]-[22]. These conditions can degrade detection performance, as the model might miss people in low-contrast or foggy scenes. Furthermore, the camera placement is constrained by the terrain and infrastructure, potentially leading to occlusions (e.g. tourists blocked from view by others or by objects) and reduced accuracy for very distant individuals. These constraints are also taken into account in designing and evaluating the system.

This study focuses on creating a capacity monitoring system for tourist attractions that utilizes real-time head count through the YOLOv8 algorithm. This study also seeks to address challenges by creating a crowd count and detection system that incorporates automatic data annotation, enhances accuracy through the addition of diverse dataset models, and achieves robustness and generalization across various datasets [12], [23]. This system aims to aid the managers of Kawah Putih Tourism Forest in effectively and accurately monitoring and regulating visitor numbers, thereby fostering a safer, more comfortable, and well-managed tourism experience.

The main contribution of this study is to integrate the head counting method with YOLOv8 to improve the accuracy of detecting individuals in dynamic environments, even under changing lighting conditions and high visitor density. This study addresses the challenges in crowd detection by utilizing automatic data annotation, enhancing accuracy with a broader range of datasets, and strengthening the model's robustness across different environmental scenarios. In contrast to earlier investigations that typically concentrate on single detection under optimal circumstances, this study introduces a system capable of functioning consistently across diverse real-world environmental settings. This study introduces a significant innovation by employing Centroid Tracker to address the challenge of double counting, a common issue arising from individuals moving within overlapping frames. By incorporating this approach, the system can monitor and tally individuals with greater precision, thus preventing inaccuracies caused by detection overlap. This study not only offers solutions for tourist attractions but also holds promise for broader applications, such as monitoring human traffic in public spaces, managing queues in shopping centers, and optimizing capacity in open areas. The proposed approach enhances crowd management effectiveness while facilitating data-driven strategic decision-making, including visit scheduling, tourist route optimization, and the implementation of real-time capacity restriction policies.

Overall, this study hypothesizes that a YOLOv8-based head detection system with centroid tracking can provide accurate and real-time visitor counts at Kawah Putih while preserving individual privacy, thus offering a viable tool for managing overtourism. The research problem addressed in this work is how to effectively monitor and quantify tourist crowds in challenging outdoor environments – characterized by high density, variable weather, and limited infrastructure – using modern computer vision. By testing this system at Kawah Putih, it aims to determine to what extent such an AI-driven approach can improve current crowd monitoring methods and support sustainable tourism management at an over touristed natural attraction. This study aims to fill the gap between theoretical carrying capacity assessment and practical on-site visitor monitoring, by providing a smart, ethical, adaptable monitoring and technological solution to help local governments measure and mitigate the impacts of overtourism in real-time [6].

In the next chapter, specifically chapter two, the Research and Methodology section will provide a more detailed explanation of the system architecture, the methods employed, and the test scenarios implemented in Kawah Putih tourism area. In chapter three, the assessment of system performance derived from test results will be detailed, covering aspects such as accuracy levels, detection efficiency, and the challenges encountered in various environmental conditions. This chapter will emphasize the advantages of the developed system in assisting tourism area managers with visitor control, ultimately enhancing capacity management, security, and the comfort of tourists. Ultimately, the Conclusion will be outlined in chapter 4.

II. RESEARCH METHODOLOGY

This section provides a comprehensive overview of the methodology employed to attain the intended outcomes. This study employs an experimental approach utilizing advanced techniques in Computer Vision and Deep Learning. It encompasses several stages, including an extensive literature review, the design of a system leveraging YOLOv8 and Centroid Tracker, the collection of image and video data

across diverse scenarios, the training of models with annotated datasets, and the testing and evaluation of system performance. The assessment was conducted utilizing a Confusion Matrix, incorporating metrics like accuracy, precision, recall, and F1-score across different test scenarios to gauge the system's efficacy in identifying and quantifying the number of individuals. The analysis results serve to pinpoint the strengths and weaknesses of the system, leading to the formulation of recommendations for future development in the discussion and conclusion sections.

A. Literature Review

This section provides a comprehensive overview of the methodology employed to attain the intended outcomes. This study employs an experimental approach utilizing Computer Vision and Deep Learning. It encompasses several stages, including an extensive literature review, system design utilizing YOLOv8 and Centroid Tracker, collection of image and video data across diverse scenarios, training of models with annotated datasets, and thorough testing and evaluation of system performance. The assessment was conducted utilizing a Confusion Matrix, incorporating metrics like accuracy, precision, recall, and F1-score across different test scenarios to gauge the system's efficacy in identifying and quantifying the number of individuals. The analysis results are utilized to pinpoint the strengths and weaknesses of the system, leading to the formulation of recommendations for future development in the discussion and conclusion sections [12].

The increasing popularity and advancement of computer vision led to its broader implementation in enhancing various facets of life. Many experts are interested in developing this technology because it has advantages in terms of speed efficiency. One of the initiatives focuses on the tourism sector, aiming to track visitor numbers in real time, thereby reducing the manual labor that is often time-consuming and energy-intensive [24]. The system operates efficiently with just a camera and a PC as its primary processor, eliminating the necessity for extensive manual effort [25].

In object detection, the YOLOv8 deep learning model is utilized within computer vision, demonstrating its effectiveness for high-speed real-time object detection. A study emphasizes the use of object detection technology and distance measurement to enhance the mobility of blind individuals, employing YOLOv8 for object detection and OpenCV for measuring distances. The study demonstrated that YOLOv8 achieved a detection accuracy of 94.2%, surpassing the 92.5% accuracy of YOLOv5. The distance measurement technique utilizing OpenCV exhibits an average error rate of 3.15% across a range of objects, including cars, doors, chairs, trees, humans, and motorbikes. This study pertains to our investigation utilizing YOLOv8 deep learning for object detection, showcasing its diverse models that exhibit competitive inference performance, making it appropriate for scenarios with differing computational speeds. Object detection identifies the head of the person in the image and highlights it with a bounding box when accurately detected. Utilizing OpenCV to compute the distance between the bounding boxes created, ensuring that undetected objects are not mistakenly counted more than once [26]-[29].

A recent study introduced the YOLOv8 deep learning algorithm designed to effectively count individuals in both still images and dynamic videos. The scenario employed in this study featured a video lasting about one minute, depicting numerous individuals entering and exiting a specific location. Following the implementation of the YOLOv8 deep learning algorithm, the initial step involved segmenting the video's area of interest to enable the system to compute the movement of individuals entering and exiting the location. The testing outcomes revealed that the model successfully identified every object in each frame with complete clarity in the results. The model effectively identified and quantified individuals accurately utilizing YOLOv8 [30]. The suggested system aligns with the one that will be introduced for counting individuals utilizing YOLOv8 deep learning. The YOLOv8 algorithm has undergone testing in multiple studies conducted earlier [31], [32]. His study utilized YOLOv8 to identify human head objects amidst different crowd densities in a tourist location, delivering real-time results.

Prior studies have highlighted the difficulties encountered by the crowded detection system. When individuals are in close proximity, overlapping objects can lead to frequent errors in the detection model's ability to differentiate between them. Inconsistent lighting due to its location in an outdoor environment. Non-ideal camera angle placement poses a challenge for those in the field. Real-time systems necessitate hardware with significant computing capabilities, posing challenges for implementation on power-efficient devices [33], [34].

The establishment of a crowd detection system in tourist attractions is a significant consideration for enhancing visitor capacity management. Numerous studies highlighted in the literature review have demonstrated the effectiveness of the YOLOv8-based crowd detection system [35]. Nonetheless, numerous innovations remain unexamined. Few studies have focused on creating systems capable of operating autonomously with energy-efficient devices [36]-[38]. Some individuals in the field do not discuss system integration for real-time monitoring. There remain limited implementations in tourist destinations characterized by dynamic settings. This literature indicates that the YOLOv8-based crowd detection system holds significant promise for application and advancement in the management of visitor crowds. Nonetheless, creating a system that can function autonomously and integrate with the Open CV interface necessitates additional investigation. This study seeks to address this gap by creating a real-time crowd detection system suitable for use in tourist attractions.

Therefore, the table below offers a comparison of various approaches in crowded counting, highlighting their respective advantages and disadvantages. Table I indicates that the proposed system delivers precise and rapid calculations while enhancing privacy. Table I presents a comparative analysis of various crowd counting methods along with their supporting technologies, benefits, and limitations. Manual counting approaches, such as observation and ticket recording, are easy to implement and require no

advanced technology [39], [40] however, they lack real-time capability and are prone to human error. Sensor-based systems using RFID or infrared can provide accurate results in controlled environments without relying on cameras [40]-[42], but they are costly and generally ineffective for monitoring large, dynamic outdoor areas. Methods based on density maps—such as CSRNet and SCNN—leverage CNN-based density estimation to provide effective crowd size approximation without needing bounding boxes [43], [44]. Despite their efficiency, these methods cannot count individuals directly, limiting their applicability in scenarios where exact headcounts are required. Object detection techniques like Faster R-CNN, YOLO, and SSD are capable of real-time individual detection using bounding boxes [45], but their performance is often hindered in high-density scenes due to occlusion and overlapping objects. The proposed method in this study, which integrates YOLOv8 with head counting and Centroid Tracking, addresses these gaps by providing a real-time, accurate, and privacy-conscious solution. Head-based detection reduces privacy risks compared to full-body detection. However, it requires a well-annotated, domain-specific dataset to improve detection performance under diverse environmental conditions.

Meanwhile, Table II presents the comparison of the YOLO method in crowd application findings indicating that the implementation of YOLOv8 is the quickest and most precise for the proposed system, which is fundamentally grounded in real-time performance. Table X presents a comparative analysis of several deep learning-based object detection models in terms of detection speed (FPS), accuracy (mAP50), strengths, and limitations. Faster R-CNN [46] offers the highest accuracy at 96%, making it ideal for applications requiring precise detection. However, its slow processing speed of 5–10 FPS renders it unsuitable for real-time scenarios. On the other hand, SSD (Single Shot MultiBox Detector) [47] achieves faster detection speeds ranging from 15 to 25 FPS, making it more applicable for real-time use, though its accuracy is lower at 88% compared to Faster R-CNN. YOLOv5 [48] presents a balanced approach with 50–60 FPS and an accuracy of 92.5%, offering a good trade-off between speed and precision. However, it still lags behind in processing speed when compared to its successor. The proposed method using YOLOv8 demonstrates as shown in TABLE I, might demonstrates the best overall performance and making it highly suitable for real-time applications such as visitor monitoring and crowd detection. Nevertheless, it requires a well-structured and task-specific dataset, particularly for head detection, to ensure optimal performance.

Table III indicates that employing head counting may prove to be more efficient in delivering insights on crowded situations, though it remains vulnerable to variations in camera angles. Nonetheless, this system is capable of preserving privacy as it does not recognize the identity of the visitor's face or body.

B. Architecture System

This section will delve into the system architecture crafted to track the real-time visitor count at tourist attractions. This system combines advanced computer vision technology based on YOLOv8 with centroid tracker algorithms to effectively detect and count individuals within a specified area. This approach allows for the optimization of visitor capacity management, enhancement of tourist experience, and assurance of security in crowded locations within the Kawah Putih tourist forest.

TABLE I. THE COMPARISON OF CROWD METHODS

Method	Technology	Advantages	Disadvantages
Manual Counting[39], [40]	Observation & ticket recording	Easy to implement, does not require advanced technology	Not real-time, prone to human error
Sensor-based (RFID/Infrared) [40]-[42]	Motion sensors, RFID, IoT	Accurate in specific environments, works without a camera	Expensive, ineffective for large outdoor areas
Density Map (CSRNet, SCNN)[43], [44]	CNN + Density Estimation	Accurate for crowd estimation, does not require bounding boxes	Cannot count individuals, only estimates crowd density
Object Detection (Faster R-CNN, YOLO, SSD)[45]	Deep Learning + Bounding Box	Real-time, directly detects individuals	Struggles with occlusion (overlapping objects)
YOLOv8 + Head Counting (Proposed Method)	YOLOv8 + Centroid Tracker	Accurate, fast, and better privacy protection compared to body	Requires specialized head detection dataset for improved accuracy

TABLE II. YOLO COMPARISON METHODS

Model	Speed (FPS)	Accuracy (mAP50)	Advantages	Disadvantages
Faster RCNN [46]	5-10 FPS	96%	High accuracy	Slow, not suitable for real-time applications
SSD (Single Shot MultiBox Detector)[47]	15-25 FPS	88%	Fast, suitable for real-time detection	Less accurate than Faster R-CNN
YOLOv5 [48]	50-60 FPS	92.5%	Balanced between accuracy and speed	Slower than YOLOv8
YOLOv8 (Proposed Method)	70-90 FPS	94.2%	Fastest and most accurate for real-time detection	Requires a specialized dataset for optimal performance

TABLE III. HEAD COUNTING VERSUS BODY COUNTING

Method	Advantages	Disadvantages	Privacy
Body Counting	More accurate in low-density conditions	Struggles in crowded conditions, overlapping objects	Poor, as it detects the entire body
Head Counting (Proposed Method)	More effective in crowded conditions, computationally lighter	Sensitive to camera angle positioning	Better, as it does not detect facial identity or full body

The block diagram illustrated in Fig. 2 represents a system architecture that comprises three primary components. The input layer utilizes the camera as the primary sensor, capturing real-time video from the tourist area. In the processing layer, the acquired video data is transformed into video streaming, with adjustments made to the frames per second to align with the system's processing capabilities. Processing improved with acceleration based on CUDA 12.6 to increase the performance of GPU-based computing. The gear used consists of an Nvidia GTX 1650 GPU and an Intel i5-11300h CPU with 16GB of RAM [38], enhancing the execution speed of the YOLOv8 model for detecting individuals within the frame. Furthermore, the centroid tracker algorithm is utilized to monitor individual movements and tally the number of visitors in each video frame [49]. Subsequently, the output layer serves as the final stage where, upon completion of processing, the detected number of visitors is presented through OpenCV.

All devices are interconnected through physical cable connections to guarantee the transmission of data. The decision to opt for a cable connection instead of a wireless one is influenced by the environmental factors present in Kawah Putih, characterized by dense fog, towering vegetation, and irregular geographical features. The use of cables enhances the system's operational reliability and provides resistance to environmental interference, thereby ensuring dependable data communication.

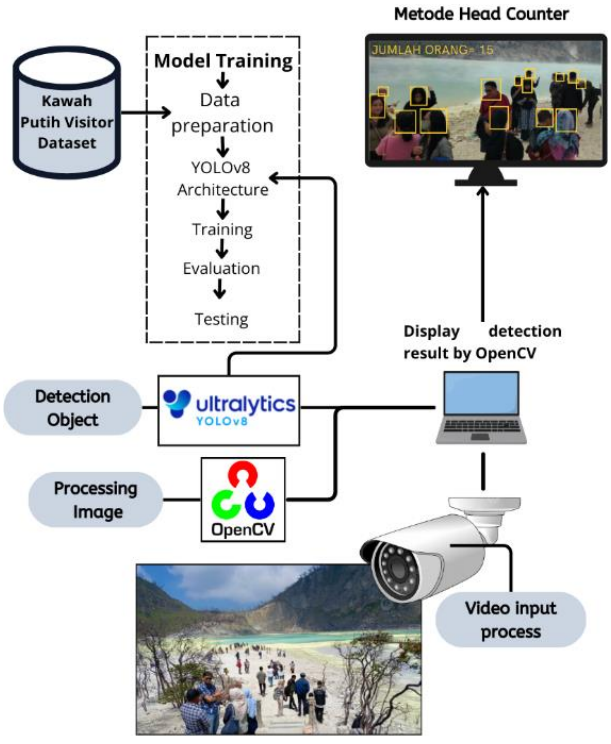

Fig. 2. Diagram block of proposed system

The design of the crowd detection system depicted in Fig. 2 demonstrates the continuous video capture by the camera, which operates at an adjusted FPS [50]. The frame rate significantly influences the quantity of frames handled each second by OpenCV. YOLOv8 handles each frame on its own, meaning that if the frames per second exceed the model's processing capability, there may be instances of frame drops, resulting in some frames being overlooked and not processed. To prevent data loss, the frame rate is set between 10-30 FPS [51], striking a balance between processing speed and detection accuracy. Once the videos have been processed, individuals within the crowd will be identified utilizing the identical dataset for every test scenario. The dataset includes a variety of lighting and weather conditions, including sunny, cloudy, and foggy weather in the Kawah Putih tourist area. These variations were used to test the robustness of the model in various real environmental situations. When YOLOv8 identifies a person, OpenCV will outline the individual's head with a blue bounding box, along with a confidence score that reflects the model's certainty in the detection [52].

The Centroid Tracker is employed to monitor the movement of individuals across successive frames. This algorithm establishes a threshold for movement distance, ensuring that a single moving individual is not counted multiple times across different detections [53]. In the final stage, the system calculates the number of bounding boxes formed to estimate the total number of visitors in the Kawah Putih tourist area. The findings from this detection are subsequently presented in the OpenCV visual interface, enabling managers to observe the visitor count in real-time and with precision [54].

C. Flowchart System

The system flowchart is presented in Fig. 3, depicting the comprehensive functionality of the system in alignment with the employed algorithm. This model is crafted to execute crowd counting through the re-identification of individuals across video frames [55]. The system is composed of multiple interconnected components that collaborate to produce precise outcomes.

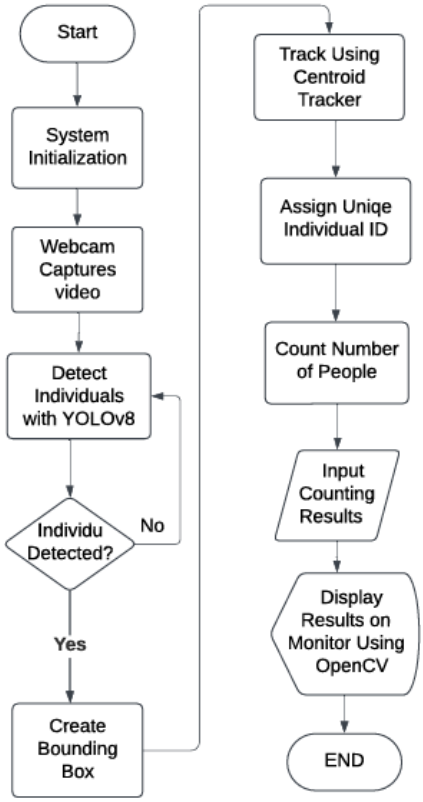

Fig. 3. Flowchart of the proposed system

Fig. 3 illustrates the complete process of the crowded detection system from beginning to end. The initial phase of the system is activated, leading to the initialization of both the software and hardware components involved. During this process, parameters including video resolution, the detection model employed, and tracking configuration are established to ensure optimal system operation. The webcam captures video through streaming, and this video is transmitted to the system as image frames for the individual detection process utilizing YOLOv8. The YOLOv8 object detection model is utilized for each frame to detect the presence of individuals [56]. Upon detection of an individual by YOLOv8, a bounding box is created around the head of each person, proceeding to the subsequent stage. If no individuals are identified, the system will revert to the video capture phase to acquire the subsequent frame.

The designed bounding box serves to indicate the individual's position in the video and acts as a reference for the subsequent tracking phase. In the subsequent phase, the identified object will be monitored using the centroid tracker algorithm to ensure that it is not counted more than once. A unique ID will be assigned to each individual to differentiate them from others. This identifier is utilized to guarantee that the system can identify the same person in later video frames. The centroid tracker has been implemented, followed by YOLOv8, which will determine the quantity of bounding boxes created and the corresponding IDs, representing the current number of visitors in the crater area [57]. The Open CV display will showcase the processed video, indicating the current number of individuals in the crater and the total number of visitors to the white crater while the system operates [51]. The count of visitors will be consistently refreshed as long as the system operates and records fluctuations in visitor density patterns.

D. Ethical Considerations

Our methodology is deliberately designed to be privacy-preserving and ethically compliant. All detections and analyses focus on anonymous features – specifically, only heads are detected, and no facial recognition or personally identifying information is used or stored at any point. The bounding boxes drawn during annotation and detection do not capture facial details, and the system does not attempt to re-identify individuals across different sessions or camera views. This ensures that visitors remain essentially anonymous in the data. By restricting the computer vision task to head counting, the design aligns with privacy principles such as those outlined in GDPR (General Data Protection Regulation) [58], which prohibits identifying individuals in surveillance footage without consent.

The data gathered by this system is limited to aggregate metrics (e.g. the number of people present over time) and non-sensitive observations like movement patterns. No video recordings or images are stored long-term with identifiable information. The processing is done in real-time and only summary statistics (total counts, dwell times) are retained for management use. This system consciously avoids any form of profiling – there is no attempt to determine a visitor's identity, demographics, or behavior beyond their presence count. The ethical approach was also communicated to the park management as stakeholders to maintain transparency about the system's purpose and limitations. In summary, the method provides a privacy-aware solution for crowd monitoring, focusing solely on operational needs (crowd size and flow) without infringing on personal privacy. Therefore, it can be ensured that the deployment of AI for smart tourism is conducted responsibly, respecting individuals' rights while still providing valuable data for managing tourist areas.

E. Environmental Conditions and Challenges

Outdoor deployment introduces several environmental challenges for vision-based detection. Variations in weather and lighting can significantly affect the reliability of the head detection. In Kawah Putih resort area, fog and mist were common in the early morning, which tended to lower image contrast and obscure distant individuals. The YOLOv8 detector's confidence dropped for heads that appeared faint or partially transparent due to dense fog. Similarly, harsh glare from the sun (especially during morning hours when sunlight hits the camera at a low angle) sometimes washed-out parts of the image, making it difficult for the model to detect heads in those overexposed regions. Another challenge was heavy rain that might be occurred during intense rainfall, the disturbance from rain streaks and water droplets on the camera lens occasionally led to false detections or missed detections. This system is mitigated some of these issues by collecting training data in such conditions and by using image preprocessing such as adjusting contrast or applying filters in software when needed, but performance naturally dips in extreme weather.

Overall, the system maintained accurate detection in mild to moderate environmental conditions, but under very adverse conditions (thick fog, direct glare, downpour) the detection reliability is reduced. Future improvements could include incorporating thermal imaging or radar in such conditions, but those are beyond the scope of this research's vision-based approach.

In terms of subject dynamics, the model showed resilience to partial occlusion and rapid motion. Because only the head needs to be visible for detection, a person half-obscured by foliage or another person can still be counted as long as their head (or part of it) is exposed. This condition is observed that YOLOv8 could detect heads even when visitors wore hats or when only the upper part of the head was visible. The detector also kept up with visitors moving quickly through the frame – at 60 FPS, even a person jogging was usually detected in most frames. However, severe occlusions and overlapping groups of people remain challenging. If a visitor's head is fully behind another person or an object (e.g. a large signboard or tree branch), the model will naturally miss that detection[59].

Similarly, when tourists cluster very tightly, their heads can overlap in the camera's 2D view, effectively appearing as a single blob. In such cases, the detector might count them as one or miss some heads, leading to an undercount. Overlapping heads can also confuse the tracker. For instance, two individuals walking side by side might be detected as one combined object, or a person passing directly in front of another can momentarily merge their detections. These scenarios sometimes caused the centroid tracker to switch

IDs or generate a short-term duplicate count. This situation is addressed by fine-tuning the tracking parameters (as noted, the distance threshold) and by implementing a logic to ignore implausibly large jumps in count within one or two frames. The tracker smooths identities over a few frames, meaning if a detection vanishes for one frame and reappears, it is likely to be treated as the same person rather than a new entrant. This helped reduce flicker in the visitor count.

Nonetheless, in very dense and dynamic crowds, the system can still experience occasional ID switches or momentary miscounts. For example, in a crowd scenario where many people enter the scene at once, the tracker may not perfectly maintain individual identities, causing a person to be counted twice or two people to momentarily share one ID. Such edge cases are inherent limitations of using only vision-based sensing with a single camera view. It will quantify these effects in the evaluation step by examining instances of counting errors during peak crowd conditions. Generally, the errors were small (on the order of a few individuals) and infrequent relative to the total count, but they underscore the need for continued improvements in multi-object tracking under heavy occlusion.

F. Human Detection Using YOLOv8

YOLOv8 transforms the challenge of object detection by reframing it as a single regression problem rather than a classification task. This algorithm allows the system to analyze the image a single time to identify the objects present and their respective locations. The system [60] divides the image into a $S\ x\ S$ grid. A bounding box will be created to indicate the object's location, accompanied by a confidence score that reflects the certainty of the box containing the object and the predicted accuracy of the box's estimation [61]. In every cell of the image grid, YOLOv8 forecasts the bounding box ($B\_(i,j)$) comprising several key components. One of these metrics is the confidence score (Ci,j), which indicates the likelihood of an object's presence within a cell and the alignment of the prediction with the actual ground truth. The calculation of this score is based on the following formula (1) and (2).

$$IoU\left(B_{i,j}, GT\right) = \frac{Area\ of\ overlap}{Area\ of\ union} = \frac{|B_{i,j} \cap GT|}{|B_{i,j} \cup GT|} \quad (1)$$

$$C_{i,j} = \Pr(Object) \times IoU(B_{i,j}, GT) \quad (2)$$

The term $Pr(Object)$ represents the likelihood of an object being present in the cell, whereas Intersection over $Union\ (IoU)$ in formula (1) quantifies the extent of alignment between the predicted bounding box and the ground truth (GT) bounding box utilized in the training process [62]. Following the prediction, YOLOv8 utilizes Non-Maximal Suppression (NMS) to eliminate duplicate detections. NMS operates by choosing a bounding box that exhibits a high confidence score and subsequently comparing it with other boxes [63]. When the intersection over union (IoU) between the chosen bounding box and other bounding boxes surpasses a specified threshold (for instance, 0.5), the bounding box that has a low confidence score will be eliminated. This process guarantees the retention of only the most precise and unique detections, thereby enhancing the overall quality of predictions [64].

The head detection model was trained on a manually annotated dataset of images and video frames captured in a real outdoor tourist environment. To enhance robustness, data were collected under diverse lighting and weather conditions, including morning periods with intense sun glare, mid-day scenes with shadowed forest trails, foggy low-contrast conditions, and overcast skies. Such diversity in capture conditions helps the model generalize to varied real-world scenarios and reduces the risk of overfitting [65]. Each image was annotated with a bounding box around each visible human head. This meticulous manual annotation effort ensured high-quality labels for training, which is crucial for reliable detection performance. In this condition, despite the environmental diversity, the dataset may not fully represent all possible visitor appearances**.** The majority of subjects come from the local tourist population, so certain ethnicities, clothing styles, or body heights could be underrepresented. This bias in the training data poses a risk to generalization – as limited diversity in training datasets can lead to biased models that perform inconsistently on unseen groups. Recognizing this limitation, the system emphasizes cautious interpretation of the system's performance in contexts involving demographics or conditions not well covered by our data.

G. Person Tracking with Centroid Tracker

Centroid tracker algorithms leverage the idea of center points to monitor the movement of objects across frames in a video. Ensuring consistent object identification across frames is a crucial challenge in computer vision, and centroid tracker algorithms provide a straightforward and efficient method to accomplish this. This method operates by identifying a bounding box surrounding an individual's head, referred to as an object in the video, and subsequently determining the center point of that bounding box. The centroid of each object is determined through the following formula (3) and (4)

$$cX = \frac{x_1 + x_2}{2} \quad (3)$$

$$cY = \frac{y_1 + y_2}{2} \quad (4)$$

The Centroid Tracker serves to identify if an individual is a novel entity or one that has been tracked previously. In the absence of any previously monitored objects, each centroid will promptly receive a distinct new identifier [66]. In cases where objects have been tracked before, the system will determine the Euclidean distance between the newly identified centroid and the previously established centroid. The equation (5) used to determine the Euclidean distance.

$$d = \sqrt{(C_{x1} - C_{x2})^2 + (C_{y1} - C_{y2})^2} \quad (5)$$

If the distance falls below a set threshold, the object is deemed identical to the previously monitored object, and the ID is preserved. When the distance surpasses the established threshold, the entity is classified as a novel object and assigned a distinct unique ID. Through the examination of

centroid movement and the implementation of a distance threshold, the bounding box is refreshed to showcase the unique ID of each detected object above it. This approach guarantees that each person is counted only once and preserves consistent identifiers, regardless of any movement within the frame. After processing all objects within the frame, the algorithm verifies that only the objects currently detected are retained in the frame. If an object is not detected in the current frame, it is eliminated from the tracking list. This guarantees that only objects currently active in the frame continue to be monitored. This method additionally aids in addressing duplicate detections and enhances the precision of real-time individual tracking [67]. The distance threshold for the Centroid Tracker is determined through empirical testing with several object movement scenarios that have been performed. The optimal value is obtained from the average maximum distance of movement between frames that can still be tracked consistently, which is around 55-75 pixels.

As previously stated, this system leverages the YOLOv8 object detection model for head detection, combined with a simple Centroid Tracking algorithm for associating detections across video frames. YOLOv8 was chosen due to its state-of-the-art accuracy and speed, making it well-suited for real-time applications[68]. The model was configured to detect human heads by training on our annotated head dataset (using transfer learning from pre-trained weights). Each detected head in a frame is represented by a bounding box, and its center point (centroid) is calculated.

The Centroid Tracker then links these detections frame-to-frame by comparing the distance between centroids in consecutive frames. It is empirically determined an appropriate distance threshold for this association process by experimentation. In our tests, threshold values of 20 pixels, 40 pixels, and 60 pixels were evaluated on sample video sequences. A moderate threshold is around 55-75 was found to work best, as it minimized double-counting and ID switches compared to lower or higher values. A too-small threshold caused the tracker to lose track of fast-moving heads (splitting one person into multiple IDs), while an overly large threshold risked merging distinct individuals into one track. Ultimately, the chosen distance threshold allowed the tracker to reliably maintain each person's identity as they moved through the scene, ensuring that a single individual is not counted multiple times.

This centroid-based data association is a lightweight approach that complements YOLOv8's detections by smoothing out frame-by-frame variations. If a detection is missing in one frame (due to momentary occlusion or detection drop), the tracker will hold the recent centroid position for a few frames, effectively smoothing the trajectory and preventing immediate loss of count. This integration of detection and tracking addresses the common issue of double counting when the same person is detected in overlapping camera frames or re-enters the scene.

H. Counting and Display the Result

The Crowded Detection system utilizes Python code to tally the number of individuals present within the camera's monitoring area. The visitor count is derived from centroid mapping using the bounding box created by YOLOv8, along with an analysis of individual movement within the video frame. The system utilizes OpenCV [69] presents the processed video frames, enabling real-time display of the calculation results. In every frame, the system presents the count of individuals presently within the monitoring area, along with the cumulative total of individuals detected throughout the monitoring session [70]. This information appears on the screen in a textual format, with colors and font sizes modified for optimal readability. Furthermore, the system employs bounding boxes along with unique ID annotations to guarantee accurate tracking of each detected individual across video frames. This method enables the system to deliver precise information regarding the number of individuals present in the monitoring area, facilitating effective crowd management and informed decision-making in the Kawah Putih tourist area.

I. Analysis Techniques

This study uses the Confusion Matrix method for data analysis to assess the performance of the Crowded Detection system. This technique is frequently employed in assessing classification models to juxtapose system predictions with actual outcomes [71], [72]. Key measures, including accuracy, precision, recall, and F1-score, can be derived from the Confusion Matrix to evaluate the system's efficacy in identifying persons under diverse testing settings. The Confusion Matrix categorizes detection results into the classifications such as True Positive (TP) defines the quantity of individuals accurately identified by the system, False Positive (FP) that denotes the quantity of entities identified by the system that are not genuine individuals (false positives). Next, False Negative (FN) that indicate the quantity of individuals present in the surveillance region who remain undetected by the system.

The Accuracy parameter reflects the system's effectiveness in correctly identifying individuals, including both the detection of their presence and the confirmation of their absence in a given area. The calculation of accuracy is determined by the formula (6). The Precision parameter serves as a valuable metric for assessing the ratio of accurate positive identifications relative to the total number of positive identifications. A greater Precision value corresponds to a reduced False Positive value. The system demonstrates a strong ability to differentiate between individuals and non-individual objects. The calculation for Precision is outlined in formula (7). In the meantime, the Recall parameter in formula (8) is assessed to evaluate the system's effectiveness in identifying all individuals within a specified area. A high recall value signifies that the system successfully identifies all relevant objects that need to be detected.

$$Accuracy = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP} + \mathrm{FN}} \tag{6}$$

$$Precision = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{7}$$

$$Recall = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{8}$$

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (9)$$

In the high-density scenario, occlusion and overlapping challenges between head objects arise. This causes the system to experience a trade-off between precision and recall, which has a direct impact on the overall performance. For example, when the model is optimized for high recall, the system tends to capture more objects, but risks detecting false positives, resulting in decreased precision. Conversely, if optimized for precision, some objects go undetected (false negatives), ultimately lowering recall. In this context, accuracy does not adequately reflect performance, as it does not consider the distribution of errors. Therefore, the F1-Score metric in formula (9) is used to evaluate the balance of precision and recall and provide a more realistic picture of performance.

III. RESULTS AND DISCUSSION

This chapter provides an in-depth examination of testing and the analysis of results. Evaluation is conducted to assess the system's effectiveness in identifying, monitoring, and quantifying individuals in real-time within a changing environment. The trial protocol encompasses 11 test scenarios, which include defining the monitoring area, methods for camera placement, and various measured parameters like detection accuracy, detection speed, and the stability of object tracking. The results are analyzed through a confusion matrix, employing metrics like precision, recall, and F1-Score to assess the system's effectiveness. The test results serve as a foundation for pinpointing possible enhancements and refining the system prior to its implementation in a real-world setting within a tourist area. Prior to executing the test, it is essential to establish the experimental setup and identify the dataset.

A. Experiment Setup

The software utilized for model training in this test comprises the Windows 11 operating system, enhanced with CUDA 12.6-based acceleration to boost GPU-based computing performance. The hardware utilized includes an Intel i5-11300h processor paired with 16GB of RAM and an Nvidia GTX 1650 GPU, facilitating the effective processing of deep learning models. The device utilized in the experiment is the Logitech Webcam C270. The development and implementation of the model were executed utilizing Python 3.10.4.

B. Dataset

The dataset utilized in this study is “Survey2,” developed by the Center of Excellence Smart Technology and Applied Science, comprising a collection of images of visitors to the Kawah Putih tourist area. Each dataset is categorized with labels, specifically ‘human’ and ‘umbrella’, comprising 425 images and a total of 5,007 annotations. When developing the image dataset for individuals, annotations are categorized into two distinct groups: ‘human’ and ‘umbrella’, particularly in instances where an umbrella is being utilized. This method guarantees that the machine learning model is capable of identifying and tallying each individual precisely, despite the presence of objects like umbrellas that may obscure portions of their bodies [48]. Items with umbrellas are labeled individually due to their partial occlusion, which enhances precision in crowd counting.

Furthermore, objects are annotated individually as explained in the previous subsection to guarantee that the model can effectively manage different degrees of object occlusion. Consequently, people adorned with head coverings like hats, hoodies, headscarves, and caps remain categorized as ‘human’ since these items do not hinder the overall visibility of the body. In contrast, umbrellas receive a distinct annotation to ensure that individuals beneath them are still identifiable, thereby minimizing the risk of incorrect detection in busy settings or outdoor scenarios [73].

This study employs YOLOv8 for the detection of human objects, highlighting that the model's accuracy is significantly influenced by the quality and specifications of the data annotation utilized in the training process. The dataset is subsequently partitioned into 70% for training, 20% for validation, and 10% for testing purposes [74]. To guarantee compliance with the YOLOv8 annotation format, Roboflow was utilized to transform the annotations and partition the dataset into three primary subsets: the training set, validation set, and testing set [75]. The illustration in Fig. 4 presents an example image sourced from the annotated dataset.

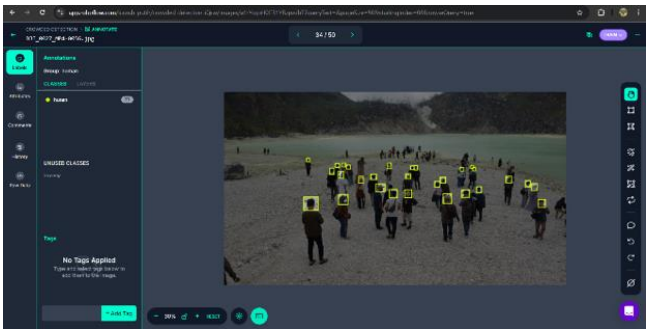

Fig. 4. Annotation process

Roboflow serves as the primary tool in the annotation conversion process, often referred to as pre-processing. This platform is instrumental in supporting the development and application of machine learning models, particularly within the realms of computer vision and object detection. This platform offers a range of tools and services designed to streamline data preparation, model training, and the implementation of models in both production and testing environments. Data preparation plays a crucial role in enhancing the precision of object identification in images or video recordings. A compilation of visuals and recordings from tourist destinations is gathered and labeled. The primary emphasis of the annotation is on individuals, whether they are walking solo, in groups, or utilizing umbrellas.

Following the object annotation phase, image resizing and augmentation techniques are applied to enhance the variety of the training data. This includes adding noise, adjusting exposure, increasing brightness, introducing color aberration, applying occlusion, and modifying saturation. This technique focuses on enabling the model to identify individuals even in challenging image quality situations, fluctuating lighting, diverse camera angles, and foggy environments, which frequently occur in practical scenarios [76]. Upon completion of the training process, the model

underwent testing with the mean Average Precision (mAP) metric to verify its performance in real-world conditions, yielding a value of 0.84. This value is regarded as favorable, given that the size of the head in the crowd image is quite small, which introduces further challenges in detection.

C. Experiments

In this test of the Crowded Detection system, a total of eleven distinct scenarios were executed to assess the effectiveness of object detection in relation to the camera's position and distance from the monitored area. The test lasted for a consistent duration of an hour of system operation. The first scenario took place in an indoor setting, featuring a maximum camera distance of 10 meters, as illustrated in Fig. 5.

Fig. 5. Indoor experiment with low light intensity

In this situation, the camera is set up indoors where lighting is consistent and the room is relatively dark, alongside a light source, while the object being detected exhibits minimal movement because of the confined environment. The objective of this scenario is to examine how the system identifies individuals in optimal conditions with minimal disruption [77].

In the second scenario, testing continues in an indoor environment, but now with an increased camera distance of 20 meters, accompanied by adequate room lighting. The primary challenge in this scenario is ensuring detection accuracy when subjects are positioned at a greater distance from the camera [78], and ensure that indoor lighting does not interfere with system performance. The test system for this scenario can be seen in Fig. 6

Fig. 6. Indoor experiment with bright light intensity

Fig. 7 displays the test results for third scenario. The camera is positioned in an outdoor setting, with a maximum distance of 10 meters from the subject being observed. The camera is positioned at an elevated height to monitor the movement of people in an open space, where elements like natural light, shadows, and the presence of non-human entities such as animals or objects may influence the detection outcomes.

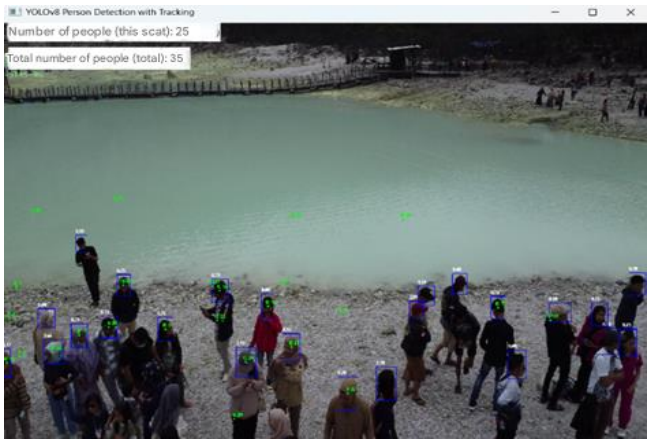

Fig. 7. 10m of camera setting at outdoor area

In the fourth scenario, the camera is positioned at a distance of 20 meters in an outdoor setting, as illustrated in Fig. 8. The system undergoes testing over extended distances to evaluate its ability to effectively detect and accurately identify individuals within a broader open environment. This includes assessments under diverse lighting conditions, accounting for fluctuations in light intensity throughout the day and into the evening.

Fig. 8. 20m of camera at outdoor scenario result

The fifth scenario presents a significant challenge regarding the placement of the camera in an outdoor environment, particularly given that the observation distance may reach up to 50 meters. The scenario illustrated in Fig. 9 aims to assess the system's ability to recognize individuals from a distance, while also examining the model's capacity to maintain accuracy as the size of the individual captured by the camera decreases. This scenario involves a critical assessment of factors including background interference, limited camera resolution, and the influence of light. This scenario facilitates a thorough assessment of the system's performance under different environmental conditions,

encompassing both controlled indoor lighting and the complex visual challenges found in open areas [79].

Fig. 9. 50 m of camera at outdoor

The sixth scenario focuses on evaluating the system's performance within a foggy crater environment, specifically analyzing the algorithm's effectiveness in identifying individuals under conditions of diminished visibility. The condition depicted in Fig. 10 poses notable challenges, as fog obscures object features and reduces detection accuracy.

Fig. 10. Foggy scenario

The seventh scenario is conducted with various camera angle variations to verify the system's ability to adjust to shifts in perspective. Adjustments in camera angles influence how objects are perceived within the frame, necessitating an assessment to identify the most effective angle that ensures maximum accuracy. Fig. 11 illustrates the camera angles captured, highlighting the white crater area located on the right side of the region

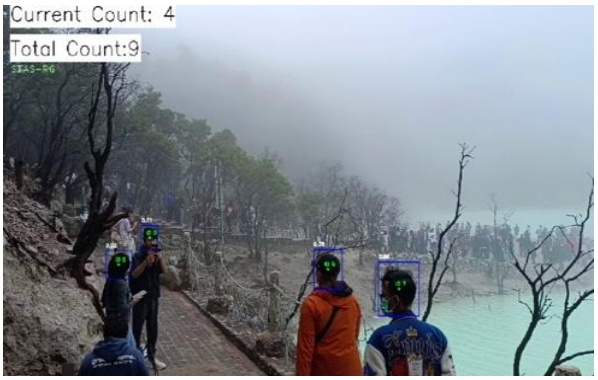

Fig. 11. Different perspective at foggy scenario

The eighth scenario simulates a crowded environment to evaluate the effectiveness of the Centroid Tracker in preventing the double counting of individuals. In this scenario, multiple entities may be in proximity or intersect, necessitating the system's capability to monitor each object without losing sight of it. In this situation, a relatively busy corner is utilized as shown in Fig. 12 to guarantee that the centroid tracker functions at its best.

Fig. 12. Crowded moving scenario

The nineth scenario show features a video that displays human figures, including not just the head but also the person holding an umbrella, as illustrated in Fig. 13. This assessment seeks to evaluate the system's capability to recognize the individual despite certain body parts being obscured by external objects, like an umbrella that may lead to occlusion.

Fig. 13. Obstacle scenario result

Furthermore, the tenth scenario involves a comparison between low-resolution cameras depicted in Fig. 14 and high-resolution cameras shown in Fig. 15, mirroring the approach taken in scenario 4, to assess the impact of image quality on object detection. The reduction in resolution may obscure object details, thus this test seeks to explore the boundaries of the system's detection abilities across different camera resolution levels.

Ultimately, the eleventh scenario evaluates the system's performance in backlight conditions, where the light source is positioned behind the object, resulting in a silhouette effect and diminishing the visibility of object details. This condition evaluates the model's ability to withstand challenges posed by low contrast and lighting constraints.

Fig. 14. Low resolution scenario result

Fig. 15. High resolution scenario result

Fig. 16. Backlight scenario

The results of these eleven scenarios will be analyzed to assess the overall system, aiming to optimize its performance across different environmental conditions and enhance the reliability of real-time object detection, focusing on precision, recall, and F1-Score metrics. This analysis reveals which scenario yields the highest accuracy in individual detection while also highlighting the system's limitations under different conditions [80]. The scenario with the highest precision and recall will be considered the best position[81], meanwhile, scenarios characterized by low accuracy will serve as a benchmark for subsequent enhancements and optimizations of the system.

D. Discussion

This system introduces a visual counter designed to monitor the number of visitors in the tourist area of Kawah Putih. The proposed system utilizes YOLO version 8 in conjunction with a centroid tracker. The system underwent testing across eleven scenarios, with the results being analyzed through a comparison with the confusion matrix and the evaluation metrics calculated. Table IV presents the test results of the crowded detection system.

TABLE IV. COMPARISON OF SCENARIO RESULTS

Scenario	Accuracy	Precision	Recall	F1-Score
1	0.9	0.92	0.96	0.94
2	0.95	0.98	0.96	0.97
3	0.86	0.97	0.88	0.92
4	0.92	0.98	0.97	0.98
5	0.63	0.9	0.675	0.78
6	0.76	0.97	0.78	0.86
7	0.8	0.95	0.78	0.85
8	0.87	0.95	0.91	0.93
9	0.83	0.97	0.84	0.9
10	0.62	0.91	0.65	0.77
11	0.89	0.92	0.96	0.94

The test results presented in Table IV provide a basis for evaluating the system. Multiple findings indicate the influence of environmental and technical factors on the performance of the crowded detection system. In the first and second scenarios, specifically indoors at distances of 10m and 20m, the system demonstrates strong performance, achieving accuracy levels of 0.9 and 0.95, respectively. Such high accuracy in indoor tests is attributed to stable lighting and uncluttered backgrounds, which make head features easier to detect. The precision and recall were likewise very high in these cases, indicating nearly all individuals were correctly identified. These findings are consistent with prior observations that controlled indoor conditions (minimal illumination changes and simpler backgrounds) tend to enhance model accuracy. The elevated precision, recall, and F1-Score suggest that the system operates effectively in indoor environments with minimal external disruption. The findings align with earlier research, indicating that indoor environments generally exhibit greater stability regarding lighting and background, which subsequently enhances the accuracy of the detection model [82].

In outdoor testing, the accuracy showed a slight decrease at a distance of 10m, dropping by 0.86, but improved at 20m, rising to 0.92. The slight drop in accuracy for outdoor clear scenarios (compared to indoors) could be due to more complex backgrounds and lighting variability; nonetheless, the system maintained robust detection capability when visibility was good, and no extreme factors were present. The metrics of precision and recall were consistently elevated, suggesting that the system maintained its capability to accurately identify individuals despite minor environmental disruptions [83]. In the fifth scenario, with the distance extended to 50m, there is a significant decline in accuracy to 0.63, recall decreases to 0.675, and the F1-score achieves only 0.78. The observed decrease suggests that, over greater distances, the model encounters challenges in accurately identifying objects. This is primarily attributed to the camera's limited resolution and the potential presence of obstacles in the outdoor setting [84].

The sixth scenario which evaluates performance in foggy conditions, demonstrates a notable effect on system accuracy, achieving a score of 0.76 a significant decline from the 0.9 level in clear conditions. This fog scenario (with heavy mist obscuring the scene) particularly challenged the model – fog introduces image distortion and reduces contrast, making it difficult for YOLOv8 to recognize head features [85], [86]. In fact, many heads that would be detected in clear air were missed in fog, likely because the blurred, low-contrast visual

cues failed to trigger the detector or matched with insufficient IoU against the true head locations. Similarly, at long-range distances**,** the system struggled when the camera was placed roughly 50 m away from the crowd, accuracy dropped to 0.63. At this range, the recall of the detector also plummeted (to 0.67). It indicating that a large fraction of people in the scene was not detected at all. The primary causes are the great distance (making each head occupy only a few pixels) and potential occlusions or background clutter at long range – the model cannot reliably distinguish or localize very small heads under those conditions. Notably, in both the foggy and 50 m scenarios, the precision remained relatively high (0.90), meaning when a head *was* detected it was usually a correct detection (few false positives). This pattern suggests the model became conservative under difficult conditions, avoiding spurious detections but missing many true heads. Such performance trade-offs underscore the challenge of maintaining high recall in low-visibility (fog) or low-resolution (distant) settings.

The seventh scenario which evaluates variations in camera angles, the observed accuracy is 0.8. The findings indicate that the system demonstrates a notable adaptability to variations in camera angles, despite a minor reduction in accuracy. In scenario 8, which simulates crowded movement, the accuracy rises to 0.87, accompanied by an F1-Score of 0.93. The result indicates that the system effectively manages individual movement within a crowd. The implementation of a centroid tracker likely enhances the system's capability to track objects while minimizing redundant calculations [52].

The nineth which evaluates the dataset involving individuals using umbrellas, the accuracy experiences a minor decline to 0.83. This suggests that the model continues to struggle with identifying individuals whose bodies are obscured by surrounding objects [87]. The tenth scenario characterized by a low resolution, demonstrates a notable decline in accuracy to 0.62 when contrasted with the fourth scenario, which features a high image resolution. The findings indicate that camera resolution is crucial for detection accuracy, as increased resolution enables the model to capture finer details. The eleventh scenario which evaluates indoor backlight conditions, the accuracy rises to 0.89, accompanied by a recall of 0.96. This outcome indicates that while backlight can influence visitability, more regulated indoor conditions enable the model to maintain strong performance [88]. The test results indicate that environmental factors, including distance, visitability, viewing angle, and camera resolution, significantly influence the performance of the crowd detection system.

The graph in Fig. 17 presents a comparison of the results from all tests. By evaluating these factors, future system development can concentrate on improving model adaptation to increase accuracy, precision, recall, and F1-Score to improve system performance[89]. In scenarios with occlusion, the IoU value tends to be lower because part of the object is covered, leading to detection errors. Similarly, in low contrast conditions, the difference between the object and the background becomes difficult to distinguish, which also leads to a decrease in the IoU value. This can be overcome by improving the detection model by increasing the variety of data provided the camera resolution and lighting are stable.

To further illustrate the system's performance variability across different environmental contexts, a heatmap was generated to visualize detection accuracy under five distinct testing scenarios as seen in Fig. 18. The heatmap clearly demonstrates that indoor environments consistently yield the highest accuracy, with values of 0.90 at 10 meters and 0.95 at 20 meters, likely due to stable lighting and minimal background clutter. Similarly, the outdoor test under clear weather conditions at a moderate range (10–20 meters) resulted in high accuracy (0.92), confirming the system's robustness when environmental visibility is optimal.

However, the heatmap also highlights notable drops in accuracy under challenging conditions. In foggy outdoor scenarios, accuracy declined significantly to 0.76, reflecting the model's difficulty in detecting low-contrast, obscured head features. The most severe performance drop was observed in the long-range (50 meter) outdoor test, where accuracy fell to 0.63. This decrease is attributed to several compounding factors: reduced object size in the frame, increased occlusions, and background complexity—all of which impair the model’s ability to reliably detect and track small, distant heads.

Overall, the heatmap emphasizes the importance of environmental context in determining the detection success rate. It provides a compelling visual summary that supports the argument for incorporating environment-aware adaptations, such as multi-camera setups or sensor fusion, to sustain accuracy across diverse real-world deployments.

To contextualize the system’s performance, a comparison with existing studies is essential shown in Table V. In a study conducted by Wang et al. [28] YOLOv5 was used for pedestrian detection in a semi-crowded outdoor environment and achieved an average accuracy of 88.3% with inference speeds around 45 FPS. While effective, the system faced limitations in high-density settings, particularly due to occlusion and overlapping subjects.

In contrast, the proposed system using YOLOv8 and head counting achieved an average accuracy of 94.2% and maintained real-time performance exceeding 60 FPS, even in complex scenarios such as overlapping heads and varied lighting. This demonstrates a notable improvement not only in detection precision but also in system responsiveness and efficiency—making it more suitable for deployment in dynamic tourism environments. The developed head detection and tracking system was evaluated in eleven different scenarios encompassing indoor and outdoor environments, varied distances, and challenging conditions. Overall, the system achieved high performance under ideal conditions – for instance, the average detection accuracy reached 94.2% (with precision 95.1% and recall 90.6%) in the test environment, and it maintained real-time processing speeds (>60 FPS) on a PC-grade GPU. These results confirm that YOLOv8 with head-focused detection provides accurate and efficient crowd counting in a controlled setting. However, the accuracy and reliability varied significantly across scenarios, highlighting the impact of environmental factors on the system’s performance.

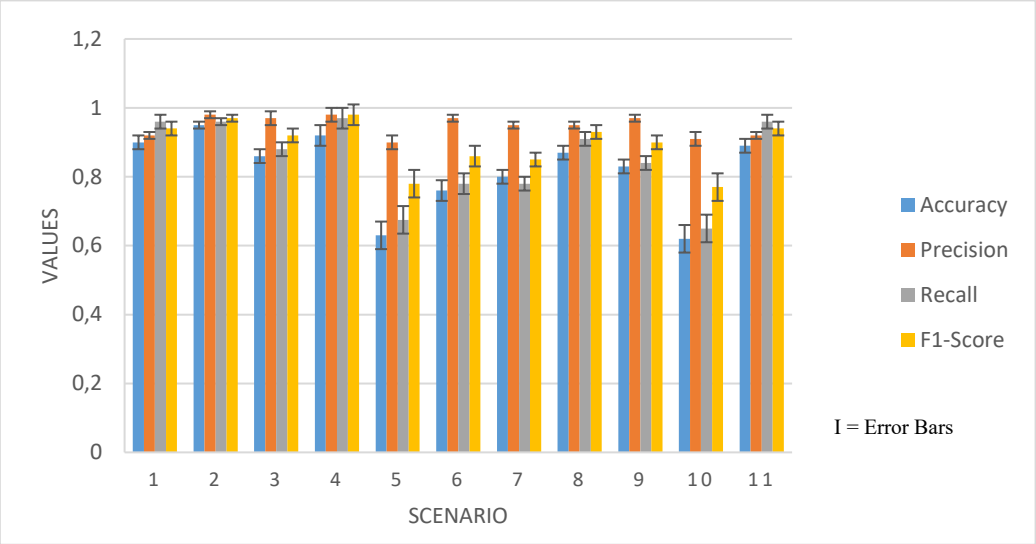

Fig. 17. Scenario comparison result

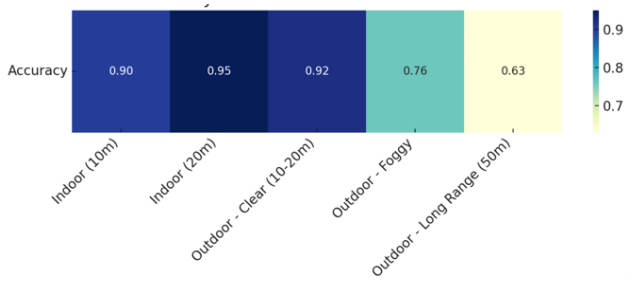

Fig. 18. The heat map of accuracy across environmental scenarios

Additionally, a study by Nguyen et al. [25] applied a conventional body-counting method using Faster R-CNN in a controlled indoor environment, reporting high precision (94%) but low frame processing speed (under 10 FPS), rendering it impractical for real-time applications in the field. Unlike that study, our system emphasizes head-based detection for privacy, and combines detection with Centroid Tracking, effectively reducing double counting errors—an issue often overlooked in earlier research.

The results of this study have several implications both practically, socially, and academically. The developed system effectively facilitates smart tourism management through the provision of real-time visitor information. This enables destination managers like Wana Wisata Kawah Putih to optimize capacity management, organize visiting schedules according to visitor density, and mitigate congestion and associated safety risks. This system enhances efficiency by decreasing reliance on manual labor for visitor count recording, thereby minimizing the potential for human error.

The head counting approach employed effectively preserves visitor privacy by ensuring that it does not identify faces or personal identities, which is significant from both social and ethical viewpoints. This aligns the system with the principles of personal data protection and supports ethical application in public spaces. Therefore, this system demonstrates both technical effectiveness and social responsibility.

The proposed system exhibits notable performance in real-time head detection and visitor counting; however, it is important to acknowledge the presence of several limitations. This system's generalizability remains limited by the scope of testing. All experiments were conducted in a single tourist location (Kawah Putih), using a dataset of visitors from that site. As a result, the model is tuned to the environmental characteristics and crowd demographics of this one location. In practice, other tourist destinations may have different background scenery, architectures, or vantage points (e.g. urban settings, beaches, museums) that could confuse a model trained mostly on forest crater images. Likewise, crowd demographics and behaviors can vary – for example, visitors in other locations might wear different styles of clothing or hats, move in different patterns, or have different group sizes. These differences may affect the head detector's accuracy if it encounters images unlike those in its training set. Therefore, further validation is needed to confirm the system works well on diverse locations and populations. The current results, while positive, are site-specific; deploying the model elsewhere might require additional training data or calibration. Expanding the dataset to include a broader range of environments (e.g. other tourist attractions with varying climates, cultural attire, and crowd densities) would improve the model's robustness and ensure it generalizes beyond Kawah Putih.

The detection accuracy diminishes as the distance increases, particularly past 30 meters. This decline is attributed to the smaller size of the detected object, complicating the YOLOv8 model's ability to accurately identify heads. The system exhibits challenges when faced with extreme lighting conditions, including backlighting, pronounced shadows, or low-light scenarios during nighttime. The positioning and orientation of the camera play a crucial role in detection performance, as less than ideal angles may lead to occlusion or partial visibility of the subject.

TABLE V. COMPARISON WITH OTHER STUDIES

Study	Methods	Accuracy	FPS	Environment	Advantages	Limitation
Wang et al. [28]	YOLOv5	88.3%	~45 FPS	Regular outdoor	Fast detection	Less effective in crowded scenes
Nguyen et al. [25]	Faster R-CNN	94%	<10 FPS	Indoor	High accuracy	Not suitable for real-time use
This study	YOLOv8 + Head + Tracker	94.2%	>60 FPS	Real-world outdoor	Accurate, real-time, privacy-aware	Accuracy decreases at long range

Another important consideration is the hardware limitation for real-world deployment. The system was tested using a PC with an Intel i5 CPU and an NVIDIA GTX 1650 GPU (as noted in the experiment setup), which easily handled the YOLOv8 model at high frame rates. Although the system functions effectively on conventional computing devices, it has not undergone comprehensive testing on low-power edge hardware like the Jetson Nano or Raspberry Pi. However, many tourism sites would prefer to run such a system on low-power edge devices (for example, an NVIDIA Jetson Nano) for portability and cost reasons. Deploying YOLOv8 on a Jetson Nano is challenging – this compact device offers substantially less processing power and no high-end GPU, so the inference speed would be much lower. In fact, the model has not yet been fully tested on such edge hardware in our study[90]. Prior experience shows that achieving real-time (>25–30 FPS) object detection on a Nano often requires careful optimization, and even then, performance may only reach a limited frame rate. For instance, running a heavy YOLO model on Jetson Nano might yield only single-digit FPS without optimizations, far from the 60 FPS observed on the PC. This gap means that additional strategies like model quantization, pruning, or using a smaller YOLOv8 variant (e.g. YOLOv8n – nano model) would be needed to meet real-time requirements on edge devices**.** Ensuring efficient operation on edge hardware is crucial for practical deployment in remote tourist sites that lack robust computing infrastructure.

A key advantage of this head-counting approach is its inherent respect for data privacy. The system deliberately avoids full-body or face recognition – it only detects and counts anonymous heads. This means it does not capture personally identifying features of individuals, aligning the system with privacy protection regulations. In jurisdictions with strict data laws (such as Europe's GDPR), video monitoring that can identify people is often considered a privacy breach unless consent is given. By focusing on heads, our method sidesteps this issue. It cannot reveal someone's identity, only their presence as part of a crowd. This significantly reduces regulatory and ethical risks, making the crowd monitoring more acceptable to the public and to authorities. The output is purely statistical (number of people and their movements) without profiling any individual. This privacy-by-design approach not only helps comply with laws but also fosters public trust, since visitors are less likely to feel that they are being personally surveilled. In summary, the choice of head-based detection provides a meaningful social benefit in that it enables effective crowd management *without* intruding on personal privacy, marrying technological capability with ethical responsibility.

Future work ought to prioritize the expansion of the dataset to encompass a broader range of environmental conditions, such as night scenes, rain, fog, and different angles. Evaluating and refining the model for embedded systems and edge devices will be essential for effective deployment. Furthermore, the incorporation of alternative tracking algorithms, such as DeepSORT or Kalman Filter, could potentially improve object tracking accuracy and minimize counting discrepancies. The integration of multiple cameras and cloud-based data analytics can enhance scalability, allow for monitoring across different locations, and support predictive crowd management in smart tourism applications.

Besides, to mitigate the performance drops observed under challenging conditions and to address the above limitations, several enhancements can be explored [74]such as, multi-camera setup covering different angles or areas can help overcome occlusions and blind spots[91], and incorporating thermal infrared cameras could allow detection of people's heat signatures in low-visibility conditions (nighttime or heavy fog). By fusing thermal imaging with the RGB camera input, the system could detect heads that are invisible or vague on the regular camera, thereby addressing scenarios with poor lighting or weather-related visibility loss. This multi-modal approach would likely increase detection recall when standard vision falters (e.g. detecting warm heads through mist or in evening hours).

Furthermore, the system can implement an adaptive resolution strategy to balance detail vs. speed. For example, the camera could capture at high resolution to detect small, distant heads (improving accuracy at long range), but the processing pipeline could downsample less critical portions of the image or use region-of-interest upscaling. Another approach is to run a lighter model or lower resolution when the crowd is sparse or close (ensuring faster inference), and switch to higher resolution or a more powerful model only when needed (e.g. when crowd density increases or people are far away). Such dynamic adjustments can maintain better accuracy without overwhelming limited hardware resources. By tuning the input size or model complexity on the fly, the system could stay near real-time on devices like Jetson Nano while still capturing important details for far or small targets.

IV. CONCLUSIONS

This research outlines the development of a real-time privacy-preserving visitor monitoring system based on YOLOv8 head detection and centroid tracking, designed to address overtourism challenges in natural tourist environments. The system demonstrated promising results in both indoor and outdoor scenarios, with high detection accuracy under optimal conditions and effective real-time

performance on mid-range hardware. By focusing solely on head counting, the system minimizes privacy risks commonly associated with facial recognition or identity tracking, offering a more ethically acceptable alternative for public surveillance. This design choice aligns with global data protection frameworks such as the GDPR, ensuring that individuals are not identifiable, and no personal data is stored.

Experiments conducted at Wana Wisata Kawah Putih reveal that the system effectively detects and counts individuals in real-time, achieving an average accuracy of 94.2%, precision of 95.1%, and recall of 90.6%, all while sustaining performance speeds exceeding 60 FPS. The incorporation of Centroid Tracking enhances tracking reliability and reduces instances of double counting, particularly in situations with overlapping objects and elevated crowd density. The implementation of automatic data annotation alongside a varied dataset has led to enhanced model generalization across different lighting and environmental conditions.

However, the practical deployment of this system in real-world environments presents notable challenges. The system's detection performance declined significantly at long-range distances (>30 meters) and in adverse weather conditions such as fog or direct glare. These limitations could severely impact the effectiveness of the solution in large, open, or sparsely monitored sites, where distant or partially visible visitors are common. In such settings, detection reliability becomes critical for informed decision-making and crowd control. Addressing these challenges will require complementary strategies, such as multi-camera networks, thermal imaging integration, or adaptive input scaling, to ensure robustness across diverse environmental contexts.

Beyond environmental constraints, ethical and social considerations must also be addressed**.** While head detection avoids identity capture, public surveillance still raises concerns regarding data governance, transparency, and informed consent, especially in regions with limited digital literacy. Moreover, the dataset used in this study, while manually curated and environmentally diverse, may not fully represent the global variation in ethnicity, attire, and head profiles. This lack of demographic diversity poses a risk of algorithmic bias, potentially leading to unequal performance across visitor populations. Ensuring fairness in detection outcomes is essential to uphold inclusivity in smart tourism technologies.

From a technical perspective, the computational demands of running YOLOv8 at >60 FPS are considerable and may be incompatible with low-power or infrastructure-constrained deployments. Although the system performed well on a GTX 1650 GPU, replicating this performance on edge platforms such as Jetson Nano or Raspberry Pi remains a challenge. To make the solution scalable, future work should prioritize hardware-specific optimization, such as model pruning, quantization, or TensorRT acceleration, and conduct comprehensive field testing on lightweight embedded devices**.** Additionally, further studies should evaluate the trade-offs between accuracy, model complexity O (n log n) tracking, and power consumption to adapt the system for sustainable, low-footprint applications in remote destinations.

Ultimately, this research contributes to the broader vision of ethical AI for public space management, where technological innovation is aligned with both environmental stewardship and human rights. By enabling data-driven visitor management without compromising individual privacy, the proposed system helps mitigate the ecological strain of overtourism while respecting social boundaries. This approach directly supports Sustainable Development Goals (SDG 11: Sustainable Cities and Communities, and SDG 15: Life on Land) by promoting responsible tourism and biodiversity protection. As a next step, the team aims to deploy and validate the system on Jetson Nano for real-world field testing, with the goal of developing a fully deployable, cost-effective, and ethically responsible solution for smart tourism infrastructures worldwide.

ACKNOWLEDGMENT

We would like to express our greatest appreciation to Telkom University for giving the resources and support that enabled this research. We also like to thank the Center of Excellence Smart Technology and Applied Sciences (STAS-RG), the Center of Excellence Smart Technology and Hospitality (STH), as well as the PT Econique research team for their crucial guidance and contributions during the project's development. Their skills and insights significantly improved the quality of our work. Thank you for your ongoing encouragement and support in furthering research in this area. This research is stated in the PKS under the number 372/LIT06/PPM-LIT/2024.

REFERENCES

[1] R. Atzori, "Destination stakeholders' perceptions of overtourism impacts, causes, and responses: The case of Big Sur, California," *Journal of Destination Marketing and Management*, vol. 17, Sep. 2020, doi: 10.1016/j.jdmm.2020.100440.

[2] E. Drápela, J. Pánek, A. Boháč, and H. Böhm, "Overtourism in the Bohemian Paradise UNESCO Global Geopark: Identifying Affected Sites Through Participatory Mapping," *Geoheritage*, vol. 17, no. 2, Jun. 2025, doi: 10.1007/s12371-025-01088-3.

[3] Y. Hassani, "Kesiapan Objek Wisata Kawah Putih Ciwidey Sambut Libur Akhir Tahun ." Accessed: Jan. 12, 2025. [Online]. Available: https://shorturl.at/jFctp.

[4] S. Liang, C. Li, H. Li, and H. Cheng, "How do you feel about crowding at destinations? An exploration based on user-generated content," *Journal of Destination Marketing and Management*, vol. 20, Jun. 2021, doi: 10.1016/j.jdmm.2021.100606.

[5] R. Prayoga, "Kunjungan ke Kawah Putih naik 44 persen pada lebaran 2023," www.antaranews.com. Accessed: Apr. 25, 2025. [Online]. Available: https://www.antaranews.com/berita/3515142/kunjungan-ke-kawah-putih-naik-44-persen-pada-lebaran-2023#:~:text=Site%20Manager%20Kawah%20Putih%20Dudung,Putih%2C%20di%20luar%20wisatawan%20mancanegara.

[6] A. Jamin and F. Rahmafitria, "Visitor Management Concept Through Carrying Capacity Analysis In Forest Recreation," vol. 5, no. 1, 2022, doi: 10.17509/jithor.v5i1.

[7] U. Suhud, M. Allan, W. C. Hoo, H. Fitrianna, and V. Noekent, "Towards Sustainable Volcano Tourism: Understanding Visit Intentions At Mount Anak Krakatau Through Destination Credibility And Environmental Motivation," *Geojournal of Tourism and Geosites*, vol. 56, no. 4, pp. 1461–1473, 2024, doi: 10.30892/gtg.56403-1317.

[8] F. C. Mihai *et al*., "Plastic pollution, waste management issues, and circular economy opportunities in rural communities," *Sustainability*, vol. 14, no. 1, p. 20, 2021, doi: 10.3390/su14010020.

[9] X. Wang, J. Liu, X. Yu, X. Chen, Y. Yu, and Z. Zhao, "People Counting with Carry-on RFID Tags," in *IEEE International Workshop on Quality of Service, IWQoS*, Institute of Electrical and Electronics Engineers Inc., 2023. doi: 10.1109/IWQoS57198.2023.10188704.

[10] H. Gomes, N. Redinha, N. Lavado, and M. Mendes, "Counting People and Bicycles in Real Time Using YOLO on Jetson Nano," *Energies (Basel)*, vol. 15, no. 23, Dec. 2022, doi: 10.3390/en15238816.

[11] S. Kapania, D. Saini, S. Goyal, N. Thakur, R. Jain, and P. Nagrath, "Multi object tracking with UAVs using deep SORT and YOLOv3 RetinaNet detection framework," in *ACM International Conference Proceeding Series*, Association for Computing Machinery, Jan. 2020. doi: 10.1145/3377283.3377284.

[12] M. Jawad Babar *et al.*, "Crowd Counting and Density Estimation using Deep Network: A Comprehensive Survey," *Journal Of Latex Class Files*, Sep. 2020.

[13] G. Curiel, K. Guerrero, D. Gómez, and D. Charris, "A Computer vision based system for human detection and automatic people counting," *Transactions on Energy Systems and Engineering Applications*, vol. 5, no. 2, Jul. 2024, doi: 10.32397/tesea.vol5.n2.624.

[14] P. Sivaprakash, M. Sankar, R. Chithambaramani, and D. Marichamy, "A Convolutional Neural Network Approach for Crowd Counting," in *Proceedings of the 4th International Conference on Smart Electronics and Communication, ICOSEC 2023*, Institute of Electrical and Electronics Engineers Inc., 2023, pp. 1515–1520. doi: 10.1109/ICOSEC58147.2023.10276183.

[15] J. Wang, M. Gao, Q. Li, H. Kim, and G. Jeon, "A Survey on Supervised, Unsupervised, and Semi-Supervised Approaches in Crowd Counting," 2024, *Tech Science Press*. doi: 10.32604/cmc.2024.058637.

[16] D. Sudharson, J. Srinithi, S. Akshara, K. Abhirami, P. Sriharshitha, and K. Priyanka, "Proactive Headcount and Suspicious Activity Detection using YOLOv8," in *Procedia Computer Science*, Elsevier B.V., 2023, pp. 61–69. doi: 10.1016/j.procs.2023.12.061.

[17] M. Hassan, F. Hussain, S. D. Khan, M. Ullah, M. Yamin, and H. Ullah, "Crowd Counting Using Deep Learning Based Head Detection," in *IS and T International Symposium on Electronic Imaging Science and Technology*, Society for Imaging Science and Technology, 2023. doi: 10.2352/EI.2023.35.9.IPAS-293.

[18] P. Shrivastav and A. K. T. Andu, "Integrated Approach for Real-time Human Counting, Tracking, and Direction Estimation using Advanced Algorithms," in *2024 15th International Conference on Computing Communication and Networking Technologies, ICCCNT 2024*, Institute of Electrical and Electronics Engineers Inc., 2024. doi: 10.1109/ICCCNT61001.2024.10726257.

[19] K. Yang *et al.*, "A Method for Quantifying Mung Bean Field Planting Layouts Using UAV Images and an Improved YOLOv8-obb Model," *Agronomy*, vol. 15, no. 1, Jan. 2025, doi: 10.3390/agronomy15010151.

[20] A. N. Alhawsawi, S. D. Khan, and F. U. Rehman, "Enhanced YOLOv8-Based Model with Context Enrichment Module for Crowd Counting in Complex Drone Imagery," *Remote Sens (Basel)*, vol. 16, no. 22, Nov. 2024, doi: 10.3390/rs16224175.

[21] R. Jofre *et al.*, "YOLOv8-based on-the-fly classifier system for pollen analysis of Guindo Santo (Eucryphia glutinosa) honey and assessment of its monoflorality," *J Agric Food Res*, vol. 19, Mar. 2025, doi: 10.1016/j.jafr.2025.101665.

[22] C. Dewi, F. Y. Bilaut, H. J. Christanto, and G. Dai, "Deep Learning for the Classification of Rice Leaf Diseases Using YOLOv8," *Mathematical Modelling of Engineering Problems*, vol. 11, no. 11, pp. 3025–3034, Nov. 2024, doi: 10.18280/mmep.111115.

[23] R. Leisha, K. J. Medows, M. M. Thiruthuvanathan, S. Ravindra Babu, P. Divakaran, and V. M. Chaturvedi, "Advanced Cyber-Physical Systems Utilizing Deep Learning for Crowd Density Detection and Public Safety," *IGI Global*, pp. 83–122, Nov. 2024, doi: 10.4018/979-8-3693-5728-6.ch004.

[24] M. Rogowski and K. Piotrowski, "Assessment and Accuracy Improvement of Pyroelectric Sensors (Eco-Counter) Based on Visitors Count in National Park. The Case: Monitoring System of Tourist Traffic in Stołowe Mountains National Park, Poland," *Environmental and Climate Technologies*, vol. 26, no. 1, pp. 182–198, Jan. 2022, doi: 10.2478/rtuect-2022-0015.

[25] H. Nguyen Viet and P. D. Phong, "Building A New Crowd-Counting Architecture," *Journal of Computer Applications in Technology, TechRxiv*, Jul. 2023, doi: 10.36227/techrxiv.23691351.v1.

[26] Erwin Syahrudin, Ema Utami, and Anggit Dwi Hartanto, "Enhanced Yolov8 with OpenCV for Blind-Friendly Object Detection and Distance Estimation," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 8, no. 2, pp. 199–207, Mar. 2024, doi: 10.29207/resti.v8i2.5529.

[27] Y. Li, T. Li, and H. Huang, "An Improved Dense Crowd Detection Algorithm Based on YOLOv5," in *ACM International Conference Proceeding Series*, Association for Computing Machinery, Oct. 2022, pp. 1713–1718. doi: 10.1145/3573428.3573731.

[28] Z. Wang, S. Yang, H. Qin, Y. Liu, and J. Ding, "CCW-YOLO: A Modified YOLOv5s Network for Pedestrian Detection in Complex Traffic Scenes," *Information (Switzerland)*, vol. 15, no. 12, Dec. 2024, doi: 10.3390/info15120762.

[29] M. A. J. Maktoof, I. T. A. Al-Attar, and I. N. Ibraheem, "Comparison YOLOv5 Family for Human Crowd Detection," *International journal of online and biomedical engineering*, vol. 19, no. 4, pp. 94–108, 2023, doi: 10.3991/ijoe.v19i04.39095.

[30] T. A. R. Shyaa and A. A. Hashim, "Enhancing real human detection and people counting using YOLOv8," in *BIO Web of Conferences*, EDP Sciences, Apr. 2024. doi: 10.1051/bioconf/20249700061.

[31] Z. Lu, L. Liao, X. Xie, and H. Yuan, "SCoralDet: Efficient real-time underwater soft coral detection with YOLO," *Ecol Inform*, vol. 85, Mar. 2025, doi: 10.1016/j.ecoinf.2024.102937.

[32] Y. L. Zeng, D. J. Guo, W. K. He, T. Zhang, and Z. T. Liu, "ARF-YOLOv8: a novel real-time object detection model for UAV-captured images detection," *J Real Time Image Process*, vol. 21, no. 4, Aug. 2024, doi: 10.1007/s11554-024-01483-z.

[33] M. Talib, A. H. Y. Al-Noori, and J. Suad, "YOLOv8-CAB: Improved YOLOv8 for Real-time Object Detection," *Karbala International Journal of Modern Science*, vol. 10, no. 1, pp. 56–68, 2024, doi: 10.33640/2405-609X.3339.

[34] H. Santosa, I. Hansen, and G. P. Kusuma, "Evaluation Of Crowd Counting Models In Term Of Prediction Performance And Computational Requirement," *Communications in Mathematical Biology and Neuroscience*, vol. 2023, 2023, doi: 10.28919/cmbn/8097.

[35] J. Nan *et al.*, "YOLOv8-LCNET: An Improved YOLOv8 Automatic Crater Detection Algorithm and Application in the Chang'e-6 Landing Area," *Sensors*, vol. 25, no. 1, p. 243, Jan. 2025, doi: 10.3390/s25010243.

[36] H. Joutsijoki and S. Mäenpää, "Crowd Counting in Action: Observations from the SURE Project," in *Smart Urban Safety and Security*, Singapore: Springer Nature Singapore, 2025, pp. 123–148. doi: 10.1007/978-981-97-2196-2_7.

[37] I. W. A. A. Wiguna, R. R. Huizen, and G. A. Pradipta, "Optimization of Vehicle Detection at Intersections Using the YOLOv5 Model," *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika (JITEKI)*, vol. 10, no. 4, pp. 885-896, 2024, doi: 10.26555/jiteki.v10i4.29309.

[38] R. Muwardi, I. P. Nugroho, K. S. Salamah, M. Yunita, R. Rahmatullah, and G. J. Chung, "Optimization of YOLOv4-Tiny Algorithm for Vehicle Detection and Vehicle Count Detection Embedded System," *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika (JITEKI)*, vol. 10, no. 3, pp. 639-648, 2024, doi: 10.26555/jiteki.v10i3.29693.

[39] A. Shokrollahi, J. A. Persson, R. Malekian, A. Sarkheyli-Hägele, and F. Karlsson, "Passive Infrared Sensor-Based Occupancy Monitoring in Smart Buildings: A Review of Methodologies and Machine Learning Approaches," *Sensors*, vol. 24, no. 5, Mar. 2024, doi: 10.3390/s24051533.

[40] K. Nalaie, V. Herasevich, L. M. Heier, B. W. Pickering, D. Diedrich, and H. Lindroth, "Clinician and Visitor Activity Patterns in an Intensive Care Unit Room: A Study to Examine How Ambient Monitoring Can Inform the Measurement of Delirium Severity and Escalation of Care," *J Imaging*, vol. 10, no. 10, Oct. 2024, doi: 10.3390/jimaging10100253.

[41] H. Wu, C. Gao, Y. Cui, and R. Wang. *Multipoint infrared laser-based detection and tracking for people counting*. Springer London, 2018.

[42] S. Comai, G. M. Stabile, E. Vavassori, M. Zerilli, A. Masciadri, and F. Salice, "Review of methods and technologies to detect, count and

identify people in indoor environments," *Internet of Things*, p. 101466, 2024, doi: 10.1016/j.iot.2024.101466.

[43] J. Gonzalez-Trejo and D. Mercado-Ravell, "Dense Crowds Detection and Surveillance with Drones using Density Maps," in *2020 International Conference on Unmanned Aircraft Systems : ICUAS'20 : June 9-12, 2020 (moved to September 1-4, 2020), Divani Caravel Hotel, Athens, Greece GR-16121*, 2020, doi: 10.1109/ICUAS48674.2020.9213886.

[44] X. Guo, M. Gao, W. Zhai, Q. Li, J. Pan, and G. Zou, "Multiscale aggregation network via smooth inverse map for crowd counting," *Multimed Tools Appl*, vol. 83, no. 22, pp. 61511–61525, Jul. 2024, doi: 10.1007/s11042-022-13664-8.

[45] R. Szczepanek, "Analysis of pedestrian activity before and during COVID-19 lockdown, using webcam time-lapse from Cracow and machine learning," *PeerJ*, vol. 8, p. e10132, Oct. 2020, doi: 10.7717/peerj.10132.

[46] M. Alruwaili *et al.*, "Deep Learning-Based YOLO Models for the Detection of People With Disabilities," *IEEE Access*, vol. 12, pp. 2543–2566, 2024, doi: 10.1109/ACCESS.2023.3347169.

[47] M. H. M. Razif *et al.*, "On Edge Crowd Traffic Counting System using Deep Learning on Jetson Nano for Smart Retail Environment," *Journal of Advanced Research in Applied Sciences and Engineering Technology*, vol. 42, no. 1, pp. 1–13, Dec. 2024, doi: 10.37934/araset.42.1.113.

[48] N. Bachir and Q. A. Memon, "Benchmarking YOLOv5 models for improved human detection in search and rescue missions," *Journal of Electronic Science and Technology*, vol. 22, no. 1, Mar. 2024, doi: 10.1016/j.jnlest.2024.100243.

[49] M. Zhang, "An Improved Fire Detection Algorithm Based on YOLOv8 Integrated with DGIConv, FourBranchAttention and GSIoU," *HighTech and Innovation Journal*, vol. 5, no. 3, pp. 677–689, Sep. 2024, doi: 10.28991/HIJ-2024-05-03-09.

[50] M. Ş. Gündüz and G. Işık, "A new YOLO-based method for real-time crowd detection from video and performance analysis of YOLO models," *Journal of Real-Time Image Processing*, vol. 20, no. 1, p. 5, 2023, doi: 10.1007/s11554-023-01276-w.

[51] D. T. Mane, S. Sangve, S. Kandhare, S. Mohole, S. Sonar, and S. Tupare, "Real-Time Vehicle Accident Recognition from Traffic Video Surveillance using YOLOV8 and OpenCV," *International Journal on Recent and Innovation Trends in Computing and Communication*, vol. 11, pp. 250–258, May 2023, doi: 10.17762/ijritcc.v11i5s.6651.

[52] S. Amaresh, R. Abhishek, P. Amal, J. Reshmi, S. K. Daniel, and S. Hashim, "Exploring the Integration of Centroid and Deep Sort Algorithms with YOLOv8 Detector for Passenger Counting in Smart Railway Systems," in *Proceedings - ICNEWS 2024: 2nd International Conference on Networking, Embedded and Wireless Systems: Wireless Technology - Building a Digital World*, 2024, doi: 10.1109/ICNEWS60873.2024.10731079.

[53] W. A. Triyanto, K. Adi, and J. E. Suseno, "Detection and Tracking of Broiler Flock Movements in the Chicken Coop using YOLO," in *E3S Web of Conferences*, 2023, doi: 10.1051/e3sconf/202344802064.

[54] S. Saha, "Traffic Monitoring System Using Machine Learning And Python OpenCV and YOLOv8," *ResearchGate*, 2024, doi: 10.13140/RG.2.2.12182.46408.

[55] S. Luo, M. Tian, and C. Xu, "Research on Tourist Flow Detection in Scenic Areas Based on Improved YOLOv8," in *IEEE Information Technology, Networking, Electronic and Automation Control Conference, ITNEC*, pp. 726–732, 2024, doi: 10.1109/ITNEC60942.2024.10733157.

[56] X. Chen, Z. Jiao, and Y. Liu, "Improved YOLOv8n based helmet wearing inspection method," *Sci Rep*, vol. 15, no. 1, p. 1945, Dec. 2025, doi: 10.1038/s41598-024-84555-1.

[57] H. Lee, D. Kang, H. Park, S. Park, D. Jeong, and J. Paik, "Real-Time Human Group Detection and Clustering in Crowded Environments Using Enhanced Multi-Object Tracking," *IEEE Access*, 2024, doi: 10.1109/ACCESS.2024.3503661.

[58] R. M. Gonçalves, M. M. da Silva, and P. R. da Cunha, "Olympus: a GDPR compliant blockchain system," *Int J Inf Secur*, vol. 23, no. 2, pp. 1021–1036, Apr. 2024, doi: 10.1007/s10207-023-00782-z.

[59] X. Zhong, G. Wang, W. Liu, Z. Wu, and Y. Deng, "Mask focal loss: a unifying framework for dense crowd counting with canonical object detection networks," *Multimed Tools Appl*, vol. 83, no. 27, pp. 70571–70593, Aug. 2024, doi: 10.1007/s11042-024-18134-x.

[60] L. Y. Xu, Y. F. Zhao, Y. H. Zhai, L. M. Huang, and C. W. Ruan, "Small Object Detection in UAV Images Based on YOLOv8n," *International Journal of Computational Intelligence Systems*, vol. 17, no. 1, Dec. 2024, doi: 10.1007/s44196-024-00632-3.

[61] S. S. Kadam, T. Jadhav, L. Patil, P. Saw, and A. Landage, "Crowd counting using yolov8 and various tracking algorithms," in *2024 OPJU International Technology Conference on Smart Computing for Innovation and Advancement in Industry 4.0, OTCON 2024*, 2024, doi: 10.1109/OTCON60325.2024.10688023.

[62] M. Kawulok and M. Maćkowski, "YOLO-Type Neural Networks in the Process of Adapting Mathematical Graphs to the Needs of the Blind," *Applied Sciences*, vol. 14, no. 24, Dec. 2024, doi: 10.3390/app142411829.

[63] A. Elaoua, M. Nadour, L. Cherroun, and A. Elasri, "Real-Time People Counting System using YOLOv8 Object Detection," in *Proceedings - 2023 2nd International Conference on Electronics, Energy and Measurement, IC2EM 2023*, 2023, doi: 10.1109/IC2EM59347.2023.10419684.

[64] F. Jing, C. Wang, J. Li, C. Yang, H. Liu, and Y. Chen, "A Dual Detection Head YOLO Model With Its Application in Wheat Ear Recognition," *International Journal of Cognitive Informatics and Natural Intelligence*, vol. 18, no. 1, 2024, doi: 10.4018/IJCINI.358013.

[65] A. Torralba and A. A. Efros, "Unbiased Look at Dataset Bias," *CVPR 2011*, pp. 1521–1528, 2011, doi: 10.1109/CVPR.2011.5995347.

[66] X. Zhou, W. Chen, and X. Wei, "Improved Field Obstacle Detection Algorithm Based on YOLOv8," *Agriculture*, vol. 14, no. 12, Dec. 2024, doi: 10.3390/agriculture14122263.

[67] S. Han, H. Ding, Z. Han, and H. Li, "Head-Dominant Enhancement With Local Count for Better Human Detection in Crowds," *IEEE Transactions on Automation Science and Engineering*, 2024, doi: 10.1109/TASE.2024.3488856.

[68] U. Sirisha, S. P. Praveen, P. N. Srinivasu, P. Barsocchi, and A. K. Bhoi, "Statistical Analysis of Design Aspects of Various YOLO-Based Deep Learning Models for Object Detection," *Springer Science and Business Media B.V*, 2023.

[69] F. Wahab, I. Ullah, A. Shah, R. A. Khan, A. Choi, and M. S. Anwar, "Design and implementation of real-time object detection system based on single-shoot detector and OpenCV," *Frontiers in psychology*, vol. 13, p. 1039645, 2022, doi: 10.3389/fpsyg.2022.1039645.

[70] D. K. D. Milton and A. R. Velraj, "Crowd Size Estimation and Detecting Social Distancing using Raspberry Pi and OpenCV," *International Journal of Electronics and Telecommunications*, vol. 69, no. 1, pp. 19–24, 2023, doi: 10.24425/ijet.2023.144326.

[71] A. N. Alhawsawi, S. D. Khan, and F. Ur Rehman, "Crowd Counting in Diverse Environments Using a Deep Routing Mechanism Informed by Crowd Density Levels," *Information*, vol. 15, no. 5, May 2024, doi: 10.3390/info15050275.

[72] Z. Tang *et al.*, "LTSCD-YOLO: A Lightweight Algorithm for Detecting Typical Satellite Components Based on Improved YOLOv8," *Remote Sens.*, vol. 16, no. 16, Aug. 2024, doi: 10.3390/rs16163101.

[73] K. Saleh, S. Szénási, and Z. Vámossy, "Occlusion Handling in Generic Object Detection: A Review," *2021 IEEE 19th World Symposium on Applied Machine Intelligence and Informatics (SAMI)*, pp. 000477-000484, 2021, doi: 10.1109/SAMI50585.2021.9378657.

[74] N. A. Pandya and N. C. Chauhan, "Multi-Camera Person Tracking: Integrating YOLOv8 with ByteTrack," *SSRG International Journal of Electrical and Electronics Engineering*, vol. 11, no. 10, pp. 53–60, Oct. 2024, doi: 10.14445/23488379/IJEEE-V11I10P106.

[75] G. I. Ortega, A. M. O. Ordas, J. F. Villaverde, R. A. Juanatas, and I. C. Juanatas, "Pedestrian-Based Adaptive Traffic Light Control Using YOLOv8," in *ICISS 2024 - Proceedings of the 7th International Conference on Information Science and Systems*, pp. 62–67, 2025, doi: 10.1145/3700706.3700717.

[76] M. M. Ali, M. S. Qaseem, M. Zeeshan, A. A. Quraishi, and A. ur Rahman, "Real-Time Crowd-Counting and Management in Sacred places using Computer Vision & ESP32 cameras," in *2023 9th International Conference on Signal Processing and Communication, ICSC 2023*, 2023, pp. 434–439. doi: 10.1109/ICSC60394.2023.10441236.

[77] M. Ş. Gündüz and G. Işık, "A new YOLO-based method for real-time crowd detection from video and performance analysis of YOLO models," *J Real Time Image Process*, vol. 20, no. 1, Feb. 2023, doi: 10.1007/s11554-023-01276-w.

[78] D. Ryan, S. Denman, C. Fookes, and S. Sridharan, "Scene invariant multi camera crowd counting," *Pattern Recognit Lett*, vol. 44, pp. 98–112, Jul. 2014, doi: 10.1016/j.patrec.2013.10.002.

[79] X. Liu, T. Shuai, and D. Liu, "Lightweight outdoor drowning detection based on improved YOLOv8," *J Real Time Image Process*, vol. 22, no. 2, Apr. 2025, doi: 10.1007/s11554-025-01638-6.

[80] A. Charef, Z. Jarir, and M. Quafafou, "Mobile Application Utilizing YOLOv8 for Real-Time Urban Traffic Data Collection," in *E3S Web of Conferences*, 2025, doi: 10.1051/e3sconf/202560100077.

[81] A. Tomar, S. Kumar, and K. K. Verma, "People Counting from Moving Camera Videos through PeopleNet Framework," *SN Comput Sci*, vol. 5, no. 8, Dec. 2024, doi: 10.1007/s42979-024-03298-y.

[82] K. Yang and A. Yilmaz, "Crowd Scene Anomaly Detection in Online Videos," in *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, pp. 443–448, 2024, doi: 10.5194/isprs-archives-XLVIII-2-2024-443-2024.

[83] M. Bhusnoor, J. Patel, A. Mehta, D. Patel, S. Sainkar, and N. Mehendale, "Investigating The Impact Of Distance On Object Detection Accuracy in Unmanned Aerial Vehicle Systems Using MobileNetV3," *Authorea Preprints*, 2023.

[84] P. Y. Leong and N. S. Ahmad, "LiDAR-Based Obstacle Avoidance With Autonomous Vehicles: A Comprehensive Review," in *IEEE Access*, vol. 12, pp. 164248-164261, 2024, doi: 10.1109/ACCESS.2024.3493238.

[85] M. Faiz, T. Ahmad, and G. Mustafa, "The Nucleus Object Detection in Foggy Weather using Deep Learning Model," *The Nucleus*, vol. 61, no. 2, pp. 117-125, 2024, doi: 10.71330/thenucleus.61.1410.

[86] H. Wang *et al.*, "YOLOv5-Fog: A Multiobjective Visual Detection Algorithm for Fog Driving Scenes Based on Improved YOLOv5," *IEEE Trans Instrum Meas*, vol. 71, 2022, doi: 10.1109/TIM.2022.3196954.

[87] S. M. Aamir, H. Ma, M. A. A. Khan, and M. Aaqib, "Real Time Object Detection in Occluded Environment with Background Cluttering Effects Using Deep Learning," *arXiv preprint arXiv:2401.00986*, 2024.

[88] W. Qi, "Object detection in high resolution optical image based on deep learning technique," *Natural Hazards Research*, vol. 2, no. 4, pp. 384–392, Dec. 2022, doi: 10.1016/j.nhres.2022.10.002.

[89] P. Gupta and U. Singh, "Evaluation of several yolo architecture versions for person detection and counting," *Multimed Tools Appl*, 2025, doi: 10.1007/s11042-025-20662-z.

[90] N. F. Abdul Hassan, A. A. Abed, and T. Y. Abdalla, "Face mask detection using deep learning on NVIDIA Jetson Nano," *International Journal of Electrical and Computer Engineering*, vol. 12, no. 5, pp. 5427–5434, Oct. 2022, doi: 10.11591/ijece.v12i5.pp5427-5434.

[91] A. Autero, M. d. M. B. Simao, and I. Karppi, *Smart Urban Safety and Security: Interdisciplinary Perspectives*. Cham, Switzerland: Springer Nature, 2025, doi: 10.1007/978-981-97-2196-2.