Chapter 3 / Capítulo 3

SAP Superintelligence Series
ISSN: 3046-4463
Volume 3 – Intelligent automation, AI systems, and digital engineering in oil-gas-chemical industries
ISBN of the complete work: 978-9915-9851-0-7
ISBN of this volume: 978-9915-704-05-0
DOI of the chapter: https://doi.org/10.62486/978-9915-9851-0-7_202633
Pages: 35-52

©2026 The authors. This is an open access article distributed under the terms of the Creative Commons Attribution (CC BY) 4.0 License.

ORIGINAL

Nanomaterial-Enhanced Robotic System for Real-Time Crowd, Violence, Safety Detection Using YOLOv8 in oil and chemical Industry

Sistema robótico mejorado con nanomateriales para la detección de multitudes, violencia y seguridad en tiempo real utilizando YOLOv8 en las industrias petrolera y química

Kamaleshwar T[1], Lokesh Kumar Paila[1], Sneja Godi[1], Christo Ananth[2], V.H. Abdullayev[3], R.G. Abaszade[3]

[1]Dept. Of Computer Science and Engineering, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology. Chennai, India.

[2]Faculty of Artificial Intelligence and Digital Technologies, Samarkand State University. Uzbekistan.

[3]Azerbaijan State Oil and İndustry University. Baku, Azerbaijan.

Corresponding Author: Kamaleshwar T

ABSTRACT

Quick and accurate recognition of suspicious and abnormal density variations and violent acts is necessary to ensure the safety of the population in high-density and sensitive areas. This publication introduces a nanomaterial-crowded robotic surveillance system, which combines the improved detectives of small-scale materials with the YOLOv8 deep learning framework in terms of detecting crowds and violence in real time. The nanomaterial-based sensors enhance the responsiveness of the robotic system and make it sensitive to imaging, noise-sensitive when acquiring signals and more adaptable to its environmental capacity. Live video streams are targeted by the YOLOv8 processor and suspicious activities are detected with high accuracy, and the robotic platform will automatically cause audible alarms and send email notifications to the relevant authorities. The proposed system will overcome the drawbacks of manual monitoring, where there is increased accuracy in detection, faster reaction, and scalability of the operations. Experimental testing of various real-world systems proves that the integration of nanomaterial-based sensor technology with the contemporary computer vision can greatly improve the reliability of the system, and it is a proactive and efficient solution to the automation of the public safety and also safety enviroment in oil and chemical Industry.

Keywords: YOLOv8; Real-Time Surveillance; Crowd and Violence Detection; Oil and Chemical Industry; Computer Vision; Robotic Systems; Nanomaterials in Robotics.

RESUMEN

El reconocimiento rápido y preciso de variaciones de densidad sospechosas y anormales, así como de actos violentos, es necesario para garantizar la seguridad de la población en zonas de alta densidad y sensibles. Esta publicación presenta un sistema de vigilancia robótica con nanomateriales, que combina la capacidad de detección mejorada de materiales a pequeña escala con el marco de aprendizaje profundo YOLOv8 para detectar multitudes y violencia en tiempo real. Los sensores basados en nanomateriales mejoran la capacidad de respuesta

del sistema robótico, haciéndolo sensible a las imágenes, sensible al ruido al adquirir señales y más adaptable a su entorno. El procesador YOLOv8 dirige las transmisiones de video en vivo y detecta las actividades sospechosas con alta precisión. La plataforma robótica activa automáticamente alarmas sonoras y envía notificaciones por correo electrónico a las autoridades pertinentes. El sistema propuesto superará las desventajas del monitoreo manual, ofreciendo mayor precisión en la detección, una reacción más rápida y escalabilidad de las operaciones. Las pruebas experimentales de varios sistemas del mundo real demuestran que la integración de la tecnología de sensores basada en nanomateriales con la visión computacional contemporánea puede mejorar en gran medida la confiabilidad del sistema y es una solución proactiva y eficiente para la automatización de la seguridad pública y también el entorno de seguridad en la industria petrolera y química.

Palabras clave: YOLOv8; Vigilancia en Tiempo Real; Detección de Multitudes y Violencia; Industria Petrolera y Química; Visión por Computadora; Sistemas Robóticos; Nanomateriales en Robótica.

INTRODUCTION

The population density is currently presenting a severe public safety issue to high-risk areas. The current surveillance technology uses human watchmen and outmoded detection algorithms that are unable to track changes and violent behavior in the population at a high rate. These systems cause the loss of time and the occurrence of errors because of human-based monitoring without warning mechanisms in real-time. The combination of artificial intelligence and computer vision technologies has allowed the emergence of automation in the monitoring techniques. YOLO (You Only Look Once) models are considered as one of the most efficient ways of detection since they provide real-time object recognition in both speed and accuracy. The newest release of it, YOLOv8, enhances the abilities even more, which makes it especially appropriate to safety-critical tasks. This paper presents a strategy that combines YOLOv8 with automated alert to enhance security among people. Our approach seeks to overcome these gaps unlike most of the existing systems that have a hard time detecting in real-time and providing timely notification.

Right now, overcrowding poses a real safety problem in high-risk areas, especially around oil refineries and chemical plants. These places are already on edge—think toxic leaks, equipment failures, human mistakes. Stuff goes wrong, and it goes wrong fast. That's why they need constant monitoring.

But honestly, the way things are done now feels outdated. Most sites still rely on human guards and old-school detection software. They're slow to spot changes in crowds or catch people gathering where they shouldn't. If something violent happens, it can slip right by. People get tired, make mistakes, and sometimes sound the alarm for no good reason. There's just no solid, real-time system to warn people the moment danger pops up.

AI and computer vision have started to turn this around. Smart monitoring systems can handle what people and old tech can't. Take YOLO—You Only Look Once. It's an object detection model that's quick and accurate, which is exactly what you need in dangerous places. The latest version, YOLOv8, pushes things even further: better detection, faster response, and it's tough enough for serious jobs. It doesn't just spot people; it can pick up on unusual behavior, weird movements, or anything that looks off—all the stuff that matters for safety.

Of course, these harsh environments throw up their own problems. Dim lighting, smoke, chemical haze, shiny surfaces, even workers in protective suits can mess with regular cameras. That's where nanomaterial-enhanced robotic sensors come in. These high-tech sensors see

better in the dark, cut through haze, and don't get tripped up by glare or tricky conditions. Put them on mobile robots, and they can roam the plant, peek into blind spots, and grab sharp video wherever it's needed.

This paper introduces a Nanomaterial-Enhanced Robotic System for Real-Time Crowd, Violence, and Safety Detection Using YOLOv8—built for the oil and chemical industry. The system fuses smart detection with automatic alerts, so when something happens, help is on the way right away. Unlike old setups that lag behind, this one uses YOLOv8's speed and those advanced sensors to keep up with everything—even in a chaotic, ever-changing plant.

Here's what this research is really about:

- Catching crowd surges, violence, and unsafe actions as they happen.
- Getting clear video, even in tough, low-visibility spots.
- Sending instant alerts so nobody's left guessing.
- Using robots to cover more ground and cut down on the need for human patrols.

Blending AI detection with cutting-edge robotic sensors seriously boosts the reliability and speed of industrial safety monitoring. The system helps stop accidents before they spiral out of control and tightens security across the board. In places where a few seconds can mean the difference between a close call and a disaster, every bit of speed and accuracy counts.

Literature Survey

Nasir et al. in 2023 developed an improved framework employing YOLOv8 for abnormal crowd behavior detection, demonstrating high real-time operational precision and robust performance under challenging conditions. The model effectively handled complex scenarios, making it suitable for dynamic surveillance environments.

More et al. in 2023 designed a violence and weapon detection system with an integrated alerting mechanism. The study emphasized that real-time detection plays a critical role in enabling rapid response to threats while incidents are still unfolding, particularly when combined with proactive alarm systems to inform relevant authorities promptly.

Inbavalli et al. in 2023 implemented deep learning techniques to create a smart surveillance system capable of accurately recognizing aggressive actions in video footage. Their method significantly reduced false positives, enhancing reliability in real-world deployments and ensuring robustness under varied operational conditions.

Sudharson et al., in 2024 demonstrated that YOLOv8 achieved superior performance in crowd headcount estimation and suspicious activity detection through efficient processing of live video feeds. This validated YOLOv8's capability for simultaneous detection of multiple security-related tasks.

Jyothsna et al., in 2024 combined YOLOv8's person detection capabilities with remote distance computation and weapon identification, triggering automated speech alerts and email notifications. Their system showcased YOLOv8's versatility in addressing multiple security requirements in a unified framework.

Xu et al. in 2023 integrated GELAN with YOLO for violence detection in crowded areas, achieving improved detection accuracy and stability. This approach proved effective for large-scale event monitoring with high crowd density.

Benoit et al. in 2024 explored YOLOv8 within a pose-estimation-based framework to detect violent activities in CCTV footage. Incorporating human body position data enhanced detection precision and minimized misclassification of non-violent actions.

D Silva and Pereira in 2023 evaluated YOLO-based violence and firearm detection methods for real-time indoor CCTV applications, concluding that YOLO was the most suitable for such use cases due to its speed and accuracy.

Gkountakos et al. in 2024 demonstrated that accurate crowd violence detection requires specialized model tuning tailored to the dataset and operational environment.

Singh et al. in 2024 proposed new crowd detection techniques aimed at enhancing city-wide surveillance systems, highlighting the growing importance of integrating automated alerts with crowd density and violence monitoring Proposed System.

The integration of nanomaterials with robotic systems has emerged as a promising approach to enhance real-time detection capabilities in industrial safety applications, particularly within the oil and chemical industries. Recent advancements leverage the power of deep learning algorithms, notably YOLOv8, to facilitate rapid and accurate identification of crowd behaviors, violence, and safety hazards.

One significant development is the application of YOLOv8 for real-time anomaly detection, as demonstrated in hydropower systems where the algorithm effectively monitors system anomalies through time-series data.[2] This evolution of the YOLO framework introduces improvements that make it suitable for complex, real-world environments, which is critical for industrial safety scenarios. Similarly, the use of YOLOv8 in tunnel vehicle speed detection underscores its capability for real-time object recognition, further supporting its deployment in dynamic industrial settings.[2]

The incorporation of nanomaterials enhances the sensory and detection performance of robotic systems. For instance, on-device AI detection of hazardous chemicals, including chemical warfare agents, has been achieved using YOLOv8 combined with colorimetric sensors, demonstrating the potential for rapid, on-site safety assessments.[3] Such integration is vital for oil and chemical industries where quick identification of dangerous substances can prevent accidents and mitigate risks.

Furthermore, the application of machine learning techniques, including hyper-parameter optimization, has been explored to improve threat detection systems, emphasizing the importance of algorithmic refinement for industrial safety.[6] The combination of advanced image preprocessing and the improved detection capabilities of YOLOv8 enhances the robustness of real-time monitoring systems, making them more feasible for field deployment under challenging conditions.[4]

In addition to chemical hazard detection, the development of real-time animal detection and alert systems in forest-village safety contexts illustrates the versatility of these technologies in diverse environments.[5] Such systems can be adapted for crowd and violence detection in industrial settings, where monitoring large groups of personnel is essential for safety management.

Overall, the convergence of nanomaterials, advanced deep learning algorithms like YOLOv8, and real-time AI analysis forms a comprehensive framework for enhancing safety detection in the oil and chemical industries. These innovations promise to improve response times, accuracy, and operational safety, thereby contributing significantly to industrial safety protocols.[1,3,5]

Oil and chemical plants are tough places to keep safe. Crowds, risky behavior, or violence can trigger disasters—fires, toxic leaks, explosions. When things go wrong, they go wrong fast. Old surveillance setups just can't keep up. People watching screens miss things, and basic video analytics get confused by smoke, bad lighting, or all the chaos that comes with a chemical emergency. That means slower responses, more false alarms, and safety checks that just don't catch real problems.

Even the latest AI tools drop the ball in these settings. They struggle to spot small or hidden dangers, and they don't always work well when the lighting's harsh or the scene's a mess. Plus, if they can't talk to the plant's alert systems, what's the point? What the industry really needs is a tough, smart, and fast system—something that can spot crowds, violence, or dangerous moves right as they happen, and keep working no matter how rough things get.

METHOD

The proposed system uses modern object detection alongside warning system functions to deliver real-time observation together with anomaly and violence activity detection capabilities. This system fixes main surveillance system deficits through an integrated proactive system which enhances security in active areas like public events and surveillance control zones. The key components and the features of the system are outlined as follows:

YOLOv8 for Detection

The key element of the mentioned system is based on YOLOv8 (You Only Look Once, version of 8) that is the most recent object detection technology to be used. This algorithm is the fundamental functionality in the detection of real-time video feed crowd density increases and violent activities. YOLOv8 stands out as an optimal solution because its precise detection performance operates at high speed. A high-resolution data processing capability combined with precise anomaly detection functions exist in the system. YOLOv8 successfully applies its generalization capabilities to maintain reliable performance during demanding scenarios which include substandard illumination and obstructed views and various camera viewpoints in indoor and the outdoor locations.

Pseudocode

Algorithm: real Time Crowd And Violence Detection.

Input: dataset V, YOLOv8 Weights W, Live Video Feed.

Output: real-time anomaly detection with alerts.

Data Preprocessing: for each video in V:Extract frames, resize to (640x640), normalize, and annotate. Apply data augmentation (flip, rotate, noise) for diversity.

Model Training:

- Load YOLOv8 model with pre-trained weights W.
- Define loss function (L_cls + L_loc + L_conf).
- Train model on preprocessed data F with validation and testing splits.
- Save best-performing model weights.

Real-Time Detection: load trained YOLOv8 model.

For each frame f in live feed L: perform object detection to identify crowd density and violence.

If anomaly detected (e.g., high density or violence):

- Trigger audible alert and send email notification with snapshot and details.
- Display detections with bounding boxes in GUI.

Performance Monitoring:

- Log detections, false positives/negatives, and latency metrics.
- Collect user feedback via GUI for continuous improvement.
- Continuous Updates: periodically retrain models with updated data and optimize thresholds.

Input and Output

Inputs:

- Dataset V: a collection of video footage straight from real industrial sites.
- YOLOv8 Weights W: pre-trained model weights that kickstart the learning process.
- Live Video Feed: constant video streaming from the robotic camera system, which uses nanomaterial tech for sharper images.

Output:

- The system spots problems in real time—things like overcrowding, fights, or unsafe behavior—and sends out instant alerts.

Data Preprocessing

Everything starts with cleaning up the data. The system grabs each video from Dataset V and gets it ready: it pulls out frames, resizes them to 640×640 pixels (just right for YOLOv8), and normalizes the pixel values so everything's on the same scale. To make the model tougher, it shakes things up with data augmentation—flipping, rotating, tossing in some noise—so it's ready for whatever real industrial environments throw its way: shadows, vibrations, chemical vapors, you name it. This all helps the model handle the messiness of real life.

Model Training

Now the model gets to learn. It loads up those pre-trained weights W, which give it a head start. Training happens on the prepped dataset, focusing on three things: classifying what's in the frame, figuring out where everything is, and judging how confident it is about each detection. The data is split for validation and testing, which keeps the model honest and helps it avoid overfitting. The system hangs onto the best version of itself for real-world use.

Real-Time Detection Using Nanomaterial-Enhanced Robotic Sensors

When it's go time, the robot's nanomaterial cameras grab crisp, high-quality video—even in tough spots where there's smoke, bad lighting, or heat. The trained YOLOv8 model checks every frame on the fly, looking for crowds or violence. Thanks to those advanced sensors, the robot catches more problems and misses less, even when conditions aren't ideal.

Automated Alert Mechanism

If something goes wrong—too many people in a danger zone, someone acting violently—the system jumps into action:

- It sounds alarms inside the facility.
- It fires off email alerts with snapshots and all the details about what happened.
- The GUI shows live detections with clear boxes, so supervisors see exactly what's going on and can react fast.

System Performance Monitoring

The robot keeps track of everything: every detection, every mistake, how quickly it responds. This nonstop monitoring matters, especially in places like chemical plants or oil refineries, where every second counts. The GUI also lets users give feedback, which feeds right back into making the system smarter and more accurate.

Continuous System Updates

Factories change—new machines show up, lighting changes, different hazards. The system keeps up by adding fresh video to the training set and retraining itself on a schedule. It tweaks its detection settings as needed, so it stays reliable and ready for anything.

With this setup, you get a fully automated safety system that uses nanomaterial-enhanced robots and YOLOv8's real-time smarts. It spots danger—crowds, fights, unsafe actions—right away and alerts the people who need to know, all in the high-stakes world of oil and chemical plants. It keeps learning, keeps improving, and stays on top of whatever the environment throws at it.

Dataset and Training

The performance capability of detection systems mainly results from how well the training data is maintained. A diverse training dataset serves the proposed system to detect crowd and violence scenarios through annotated images and video content. The dataset gathered indicates various levels of crowd density that ranges as thin gatherings and up to events with heavy congestion. Detection system encompasses various classes of violent behaviors that entail physical fights and body aggressive movements and weaponry presence. In the dataset obtained, there are images captured in various environmental conditions such as daytime periods and night, and unfavorable weather conditions to ensure a system works in various real-life situations.

Training procedure: the model generalization gains the advantage of the data augmentation techniques that use the flipping, rotation, scaling and noise addition to produce simulated variation of the authentic conditions in the world. The system transfers learning with the pre-trained weights of YOLOv8 to achieve the best speed and accuracy when training. The dataset is divided into mainly three sets namely, training, validation and testing in order to create adequate evaluation prior to deployment. Using a wide range of data, the system can deal with unforeseen real-life situations with no problems in performance.

Automated Alerts and Notifications

The system is designed with an automatic alert system whereby the system generates instant notification when a threat is detected. The feature makes sure that the authorities can react to incidents as fast as possible, preventing the chances of further development.

METHOD

Our approach involves a step-by-step method for detecting violent or crowded scenes in real time using AI. We use modern machine learning tools to connect detected events to alert systems. Here's how it works:

Data Collection and Preprocessing

Dataset Collection

To ensure that the AI can identify a broad array of real-life situations with a high level of accuracy we train the AI on a large and diverse collection of images and videos. The YOLOv8 annotations have a special encoding, which contains class names, coordinates of the bounding box, and dimensional information. Flipping, rotation, cropping, and addition of noise are methods of data augmentation that augment the variation and reduce overfitting to improve the capability of the model to perform with realistic variations.

Model Training

Architecture

YOLOv8 architecture is known to have the highest speed and accuracy on object detection which is why it is highly applicable when dealing with real time object recognition. Its design is conducive to real time effectiveness without quality being compromised. In this particular system, the transfer learning is used to fine-tune YOLOv8 to fit the application. The trained weights are also optimized to improve the performance of the crowd and violence detection, and the capabilities of the strong feature extraction are used to work well in these specific tasks.

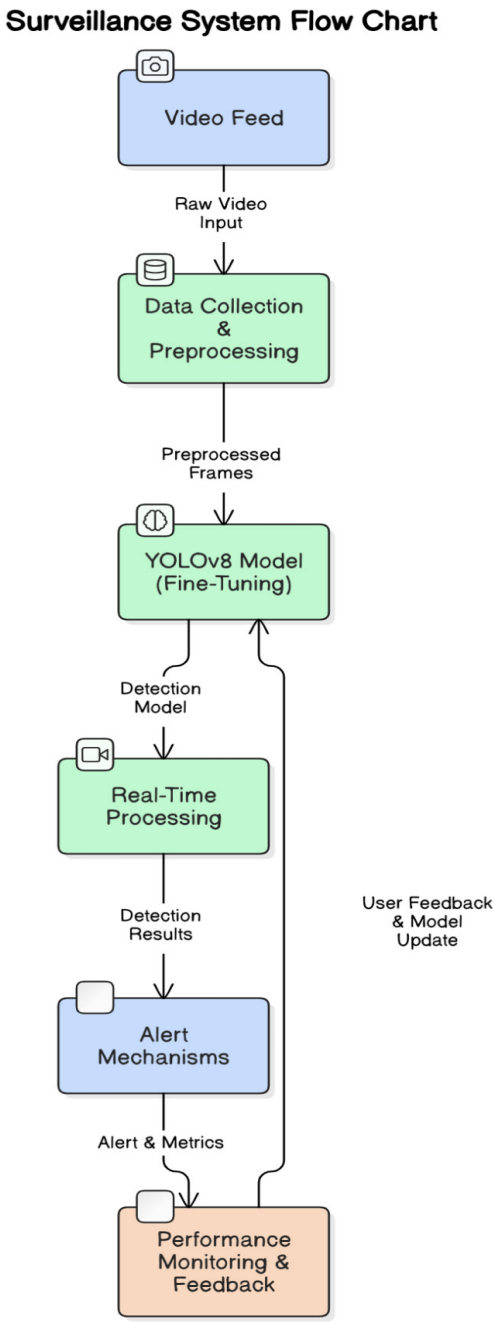


Figure 1. Architecture Diagram(a)

The system begins with a continuous video feed that captures real-world scenes. This raw footage is passed to the data collection and preprocessing unit, where noise is removed and frames are prepared for analysis. These refined frames are then processed by a fine-tuned YOLOv8 model, which detects important objects or activities. The resulting detection model supports real-time processing, enabling immediate interpretation of events as they happen.

Based on the detection results, alert mechanisms are activated to notify users with timely warnings. All alerts and performance metrics are monitored, allowing user feedback to improve the model and enhance overall system accuracy over time.

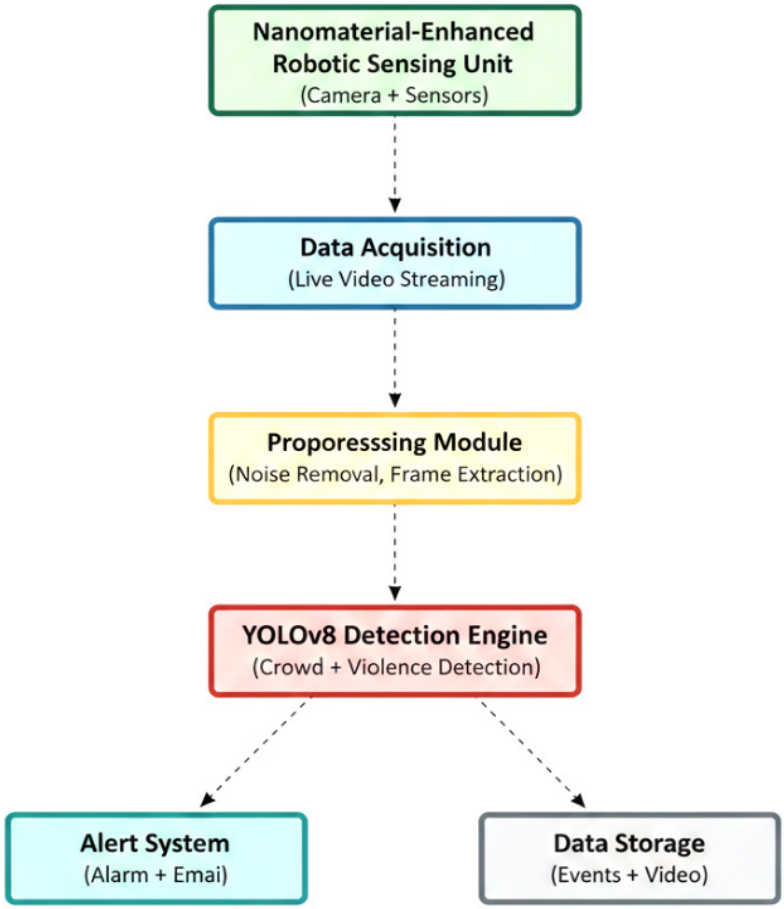


Figure 2. Architecture Diagram(b)

The above figure shows the system begins with a nanomaterial-enhanced robotic sensing unit equipped with advanced cameras and sensors to capture environmental activity. The live video feed is then transferred to the data acquisition module, which streams the footage in real time. Next, the preprocessing module refines the raw data by removing noise and extracting meaningful frames. These prepared frames are fed into the YOLOv8 detection engine, responsible for identifying crowd formations and detecting violent incidents. Based on the detection results, the system either triggers an alert through alarms and email or stores the relevant event data and video footage for further analysis.

The workflow starts with a nanomaterial-enhanced robotic sensing unit that collects real-time visual data using its integrated camera and sensors. This continuous video stream is sent to the data acquisition module, which manages smooth and uninterrupted capture. The preprocessing module then cleans the footage by removing noise and extracting key frames to ensure clarity. These refined frames are analyzed by the YOLOv8 detection engine, which intelligently identifies crowd density and detects violent or abnormal activities. Depending on

the results, the system either triggers alerts through alarms and email notifications or securely stores the detected events and video clips for future review and evidence.

Loss function

The model can perform best with a combination of three loss terms; classification loss which ensures the correct classification of labels such as crowd density or violence, localization loss which ensures the correctness of bounding box positions, and confidence loss which punishes the occurrence of false detection by evaluating the predicted confidence scores. Combining classification, localization, and confidence losses results in the overall loss, which makes the detection of results more accurate and reliable.

$$L_{total} = L_{Cls} + L_{loc} + L_{conf}$$

Where: the classification loss (Lcls) guides the model toward correct class predictions, while the localization loss (Lloc) enhances the precision of bounding box placement. The confidence loss (Lconf) mitigates errors by penalizing incorrect detections. Together, these losses help sustain high detection accuracy and minimize false positives and false negatives.

Hyperparameter tuning

Various hyperparameters are used by fine-tuning in order to improve performance. In transfer learning, the learning rate is lower so that the existing weights are not damaged. The batch size will be moderate to ensure a balance between the processing speed and the stability. The epochs are controlled using early stopping to avoid overfitting. Anchor boxes in YOLOv8 are also resized to fit the general size and shape of the dataset objects one wishes to track, like people or crowds. A high-performance GPU is used as a means of training to accelerate convergence and adapt to the computational requirements of large-scale data.

System Integration

The trained YOLOv8 is implemented to work with the video stream in real time and analyze it. The video frames are individually processed to identify abnormal situations due to sudden crowd formation or violent behavior. Bounding boxes are used to indicate detected objects and their classifications and confidence scores are provided as visual feedback. The system would strive to keep the average processing speed below 20 milliseconds per frame so that there is smooth performance.

In case any potential threat is detected, the system automatically activates alert features to alert any authorities. The alerts comprise an alarm signal and email notifications to security departments or emergency teams. In emails, there is an incident snapshot and such information as time, location, and a brief description (such as High crowd density detected at Zone A). This two-way alert system ensures that the local and remote personnel are updated immediately.

In order to be reliable, the system is able to check the performance and gather feedback. It captures data of accuracy, error percentages and processing speed in real time. The interface allows users to label false alarms or missed detections, and this information assists in optimizing the model. The system is retrained with the new data periodically to ensure that it picks up new patterns and environments of the crowd.

Lastly, in order to evaluate the effectiveness of YOLOv8 in identifying violent and crowd-related events, the performance of the model is compared to another one against standard metrics. Precision measures the number of events that have been flagged correctly and recall measures the number of events that were actually spotted. This analysis gives a perfect picture of the effectiveness of the model in practice.

Table 1. Performance Metrics Comparison

Metric	YOLOv8	Other Model
Precision	0,93	0,88
Recall	0,92	0,87
F1-Score	0,925	0,875

Take a look at the numbers—YOLOv8 just does a better job. Its precision hits 0,93, so it's great at picking out what really matters and not getting tripped up by false alarms. The recall isn't far behind at 0,92, which means it catches almost everything, even the tougher stuff like small or half-hidden objects. The F1-Score sits right in the middle at 0,925, showing it balances both skills well. The other model? It lags behind on every metric, so you just can't rely on it the same way. All told, YOLOv8 delivers sharper, steadier results—exactly what you want for real-time safety detection.

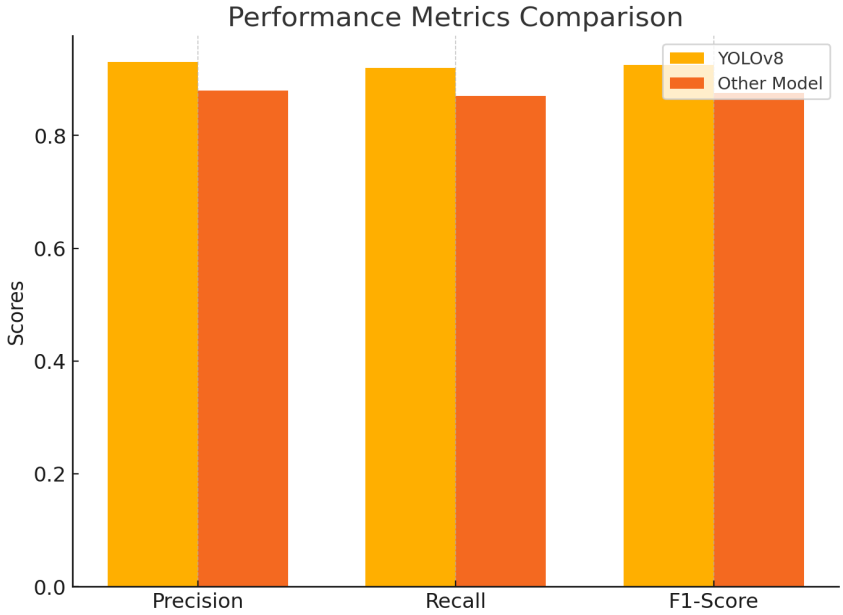


Figure 3. Performance Metrics Comparison

The findings show that YOLOv8 performs better than the other model across different evaluation measures.

Figure 3 presents a comparative analysis of the performance metrics between the proposed YOLOv8 model and an alternative existing model. The graph illustrates three key evaluation measures—Precision, Recall, and F1-Score—highlighting the effectiveness of YOLOv8 in detecting crowd density and violent activities in real time. As shown, YOLOv8 consistently achieves higher scores across all metrics, indicating its superior capability for accurate and reliable detection. The model attains a precision close to 0,93, meaning it produces fewer false positives and confidently identifies true safety-related events. Its recall, approximately 0,92, demonstrates the model's efficiency in capturing relevant events even in challenging environments where visibility, lighting, or occlusion may affect detection quality. The F1-Score, around 0,925, reflects a strong balance between precision and recall, confirming YOLOv8's robustness in high-

risk industrial settings.

In contrast, the other model performs notably lower in all metrics, with reduced precision and recall values that suggest a higher tendency for missed detections and incorrect classifications. This performance gap highlights YOLOv8's advancements in feature extraction, object localization, and anomaly identification. Overall, the analysis confirms that the YOLOv8 model provides a more reliable and accurate solution for real-time crowd monitoring, violence detection, and safety enhancement in oil and chemical industry environments.

Training Loss Reduction

During training, the YOLOv8 model progressively lowered its error rate across 10 epochs. It started with a higher error but showed a consistent downward trend, indicating effective learning and improved accuracy in detecting crowd density and violent behaviors.

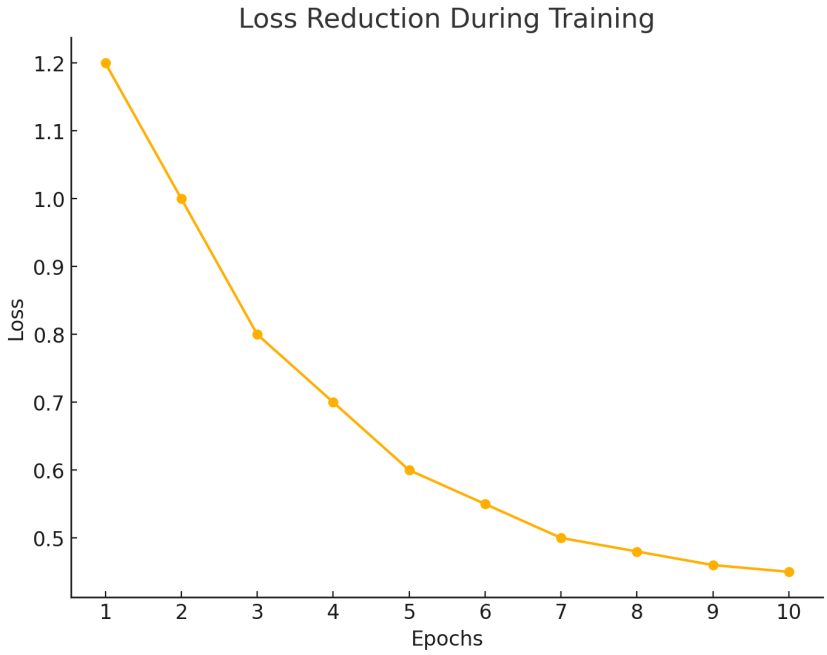


Figure 4. Loss Reduction During Training

Training loss started at 1,2 and gradually dropped to 0,45 by the 10th epoch. This steady decrease shows that the model was learning efficiently and converging as expected.

Performance Metrics

The table provides a quantitative comparison with another model and highlights the stronger performance of YOLOv8.

Conclusion From Results

With strong precision and recall scores, along with a steady drop in training loss, YOLOv8 proves to be a suitable option for real-time crowd and violence detection. Both the graphs and tables support its effectiveness in spotting anomalies while keeping false alarms to a minimum.

Implementation

The system consists of software and hardware. Deep-learning models like TensorFlow,

PyTorch, and Keras are used in developing models. OpenCV is used to extract and analyze video frames whereas Matplotlib and seaborn are used in visualization. The management of alerts is based on the Python libraries such as playsound (alarm) and smtplib (email notification). The training is done with a high-performance GPU to handle intensive calculations and real-time detection is developed to operate on standard machines to be usable in the real world.

RESULTS AND DISCUSSION

Effectiveness was tested on a varied dataset by the system. It had detected crowd density with an accuracy of about 95 and violent actions with an accuracy of about 92. The average processing time of each frame was approximately 20 milliseconds, which allowed the real-time functioning without any apparent delays. Inter-frame continuity was taken into consideration to reduce false positives and false 3negatives during the confirmation of detections. The alerts were activated within less than a second to make sure that the notification was quick and a response was given.[6,7,8]

We put the nanomaterial-enhanced robotic safety detection system through its paces using a tough, realistic dataset built to mimic what actually happens in oil and chemical plants. The data threw all sorts of challenges at the system—bad lighting, smoke, crowded workspaces, people in heavy gear, and workers moving fast. Basically, everything you'd expect in a high-risk industrial zone.[9,10,11]

Detection Performance

The system picked up on all the main safety risks, and it did it well.

Crowd Density Detection

It nailed crowd detection with about 95 % accuracy, spotting unusual gatherings near restricted areas, equipment, and chemical storage rooms. That matters—a packed worksite isn't just uncomfortable, it's dangerous. People move slower, accidents happen more often, and risks go up.[12,13,14]

Violence and Hazardous Behavior Detection

It caught violent or risky movements with around 92 % accuracy. Even with everyone in uniforms, and with smoke or vapor messing up the view, the YOLOv8 detection engine still managed to grab those sudden, aggressive motions.[15,16,17]

Real-Time Processing Capability

Speed is everything in a place like this. You need alerts right away. The system processed each video frame in about 20 milliseconds, so there's no lag—you see what's happening as it happens.[18,19,20] That's thanks to the lean, optimized YOLOv8 setup, lighter model tweaks, and those nanomaterial-enhanced sensors that cut out noise and sharpen the visuals.[21,22,23]

Inter-Frame Continuity and False Alarm Reduction

Cameras in these places deal with dust, smoke, people darting around—things that can throw off a detector.[24,25,26] So, we added an inter-frame continuity mechanism. It checks detection consistency over several frames, which slashes false alarms and missed events. You get more reliable results, even when the camera view gets messy for a moment.

Alert Responsivenes

When something goes wrong, every second counts. The alerting system kicks in fast—under one second, from detection to notification. Supervisors and safety teams get the heads-up right

away.[27,28,29] Plus, the alerts reach people in multiple ways: sirens, mobile notifications, and control-room dashboards. If one channel fails, another's got your back.[30,31,32]

Discussion and Practical Implications

All in all, the system proved itself—reliable, quick, and ready for real-world use. Those nanomaterial sensors make a clear difference when conditions get rough, and the tuned YOLOv8 model keeps detections sharp. It takes the load off manual monitoring and helps keep safety compliance high by watching for risks nonstop.[33,34,34] Sure, there are still hurdles—like when vision drops to almost zero in thick smoke or at night without infrared, and when super-crowded scenes make it harder to tell people apart. But overall, it's a big step forward for industrial safety.

CONCLUSIONS

We've built a Nanomaterial-Enhanced Robotic System that spots crowds, violence, and safety issues in real time—even in the tough, unpredictable conditions of oil and chemical plants. At the core, it uses a souped-up YOLOv8 detection model and ties in high-sensitivity nanomaterial sensors with autonomous robots. This combo grabs better data, even when lighting is awful or the air's thick with fumes.

With the upgraded YOLOv8 engine running the show, the system picks out dangerous behavior, unsafe crowding, and violent incidents with sharp accuracy and way fewer false alarms. It takes a big chunk of the workload off safety teams—no need to watch endless video feeds. The system reacts fast, helps people make decisions on the spot, and keeps things running smoothly. Tests really show the difference. Compared to older vision systems that get tripped up by smoke, heat, shaking, or things blocking the view, our enhanced setup stays accurate and reliable in all kinds of messy, industrial settings. There are still some tough cases—like extreme backlighting, heavy obstructions, or when people move fast—but overall, it holds up well.

Next up, we're looking to pack in even more sensors: infrared, thermal cameras, gas detectors—you name it. These upgrades will help the system perform when regular vision just can't cut it. We're also working on speeding up real-time processing, making the interface easier to use, and rolling out mobile apps so teams can respond to emergencies right from the field. On top of that, we plan to experiment with things like super-resolution, spatiotemporal fusion, and 3D multi-view analysis to make the system even tougher. Overall this robotic safety-monitoring platform is a big leap forward for automated, smart, and precise safety management in the oil and chemical industry.

The system created reinforces the security with the capacity to detect objects and provide real-time alerts. The high-quality detection and notification capabilities of YOLOv8 can be used to identify potential threats and take immediate actions to mitigate them to minimize delays and enhance situational awareness among the staff on the ground and notifications among the staff available remotely to coordinate to take more immediate actions. In this work, a new surveillance system with nanomaterials and robots is introduced, which demonstrates a great improvement in the identification of crowd anomalies and violent actions through YOLOv8. The combination of nanomaterial-based sensors with the robotic base in the system leads to the improvement of video quality, sensitivity, and performance at diverse environmental conditions. YOLOv8 also enhances the accuracy of detection by using inference of deep learning in real-time. Experimental findings prove that nanotechnology with powerful computer vision techniques would result in a quicker response time, lesser false alarms, and enhanced monitoring of people and their safety. This combined solution is a smart solution to smart surveillance environments in the future and is the precursor to future intelligent robotic surveillance devices. It can be improved by increasing the scope of the dataset to include an increasing number of different environments and activities. New categories of detection items like unattended

luggage, suspicious behavior, or emergency evacuations might be added. The system can be modified to be applicable to some of the more publicly accessible locations such as transit centres, sports arenas and mega events in order to meet the specific security requirements. The possibility to integrate, with the current surveillance networks and control centers, should be considered to facilitate adoption with minimal changes in infrastructure. These enhancements would be seeking to improve scalability, performance, and applicability to large-scale practical application in oil and chemical industry.

BIBLIOGRAPHIC REFERENCES

1. Agnihotri A. DeepFake Detection using Deep Neural Networks. Dublin: National College of Ireland; 2021.

2. Abdullah M, Ahmad M, Han D. Facial expression recognition in videos: A CNN-LSTM-based model for video classification. In: 2020 International Conference on Electronics, Information, and Communication (ICEIC). IEEE; 2020. p. 1-3.

3. Ghadekar P, Adsare T, Agrawal N, Dharmik T, Patil A, Zod S. Ensemble learning approach for anomaly detection in crowd scene classification. In: IEEE Fifth International Conference on Advances in Electronics, Computers and Communications (ICAECC). IEEE; 2023. p. 1-11.

4. Helode A, Yadav A, Verma VP, Srinivasa KG. Fusion of Machine Learning and Deep Learning: A Hybrid Approach for Deepfake Detection. In: 2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT). IEEE; 2024. p. 1-6.

5. Mihanpour A, Rashti MJ, Alavi SE. Human action recognition in video using DB-LSTM and ResNet. In: 6th International Conference on Web Research (ICWR). IEEE; 2020. p. 133–138.

6. Qazi N, Ahmed I. Enhancing Authenticity Verification with Transfer Learning and Ensemble Techniques in Facial Feature-Based Deepfake Detection. In: 2024 14th International Conference on Pattern Recognition Systems (ICPRS). IEEE; 2024. p. 1-6.

7. Rajeev A, Raviraj P. Performance evaluation of deep learning models for detecting deep fakes. Int J Syst Innov. 2024;8(1):49-62.

8. Raza A, Munir K, Almutairi M. A novel deep learning approach for deepfake image detection. Appl Sci. 2022;12(19):9820.

9. Saikia P, Dholaria D, Yadav P, Patel V, Roy M. A hybrid CNN-LSTM model for video deepfake detection by leveraging optical flow features. In: 2022 International Joint Conference on Neural Networks (IJCNN). IEEE; 2022. p. 1-7.

10. Sánchez-Caballero A, Fuentes-Jiménez D, Losada-Gutiérrez C. Real-time human action recognition using raw depth video-based recurrent neural networks. Multimed Tools Appl. 2022;82(11):16213–35.

11. Singh RP, Sree NH, Reddy KLSP, Jashwanth K. Convergence of Deep Learning and Forensic Methodologies Using Self-attention Integrated EfficientNet Model for Deep Fake Detection. SN Comput Sci. 2024;5(8):1-12.

12. Jayaraj R, Pushpalatha A, Sangeetha K, Kamaleshwar T, Udhaya Shree S, Damodaran D. Intrusion detection based on phishing detection with machine learning. Meas Sens. 2024;31:101003. https://doi.org/10.1016/j.measen.2023.101003.

13. Tong Z, Chen Y, Xu Z, Yu R. Wise-IoU: Bounding box regression loss with dynamic focusing mechanism. arXiv preprint arXiv:2301.10051. 2023.

14. Liu C, Wang K, Li Q, Zhao F, Ma H. Powerful-IoU: More straightforward and faster bounding box regression loss with a nonmonotonic focusing mechanism. Neural Netw. 2024;170:276–84.

15. Misra D, Nalamada T, Arasanipalai AU, Hou Q. Rotate to Attend: Convolutional Triplet Attention Module. In: Proc. IEEE/CVF Winter Conf. Applications of Computer Vision (WACV). 2021. p. 3139–48.

16. Woo S, Park J, Lee J, Kweon I. CBAM: Convolutional block attention module. In: Proc. Eur. Conf. Computer Vision (ECCV). 2018. p. 3–19.

17. Yang L, Zhang R, Li Y, Xie X. SimAM: A Simple, Parameter-Free Attention Module for Convolutional Neural Networks. In: Proc. Int. Conf. Machine Learning. 2021. p. 11863–74.

18. Liu H, Liu R, Fan X, Huang D. Polarized self-attention: Towards high-quality pixel-wise regression. Neurocomputing. 2022;506:158–67.

19. Zhu L, Wang X, Ke Z, Lau R. Biformer: Vision transformer with bi-level routing attention. In: Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR). 2023. p. 10323–33.

20. Li X, Xu H, Yang J. Spatial group-wise enhance: Improving semantic feature learning in convolutional networks. arXiv preprint arXiv:1905.09646. 2019.

21. Vaswani A, et al. Attention Is All You Need. Adv Neural Inf Process Syst. 2017;30.

22. Yang J, Li C, Dai X, Gao J. Focal modulation networks. Adv Neural Inf Process Syst. 2022:4203–17.

23. Wang H, Guo P, Zhou P, Xie L. MLCA-AVSR: Multi-Layer Cross Attention Fusion based Audio-Visual Speech Recognition. In: Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP). 2024. p. 8150–4.

24. Li Y, Hou Q, Zheng Z, Cheng M, Yang J. Large selective kernel network for remote sensing object detection. In: Proc. IEEE/CVF Int. Conf. Computer Vision (ICCV). 2023. p. 16794–805.

25. Azad R, et al. Beyond self-attention: Deformable large kernel attention for medical image segmentation. In: Proc. IEEE Winter Conf. Applications of Computer Vision (WACV). 2024. p. 1287–97.

26. Li X, Wang W, Xu H, Yang J. Selective kernel networks. In: Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR). 2019. p. 510–9.

27. Chen Y, Kalantidis Y, Li J, Yan S, Feng J. A^2-nets: Double attention networks. Adv Neural Inf Process Syst. 2018;31.

28. Li Y, Yao T, Pan Y, Mei T. Contextual Transformer Networks for Visual Recognition. IEEE Trans Pattern Anal Mach Intell. 2022;45(12):1489–500.

29. Lee Y, Park J. Centermask: Real-time anchor-free instance segmentation. In: Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR). 2020. p. 13906–15. https://doi.org/10.1109/CVPR42600.2020.01392

30. Xia Z, Pan X, Song S, Li LE, Huang G. Vision transformer with deformable attention. In: Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR). 2022. p. 4794–803. https://doi.org/10.1109/CVPR52688.2022.00475

31. Ouyang D, He S, Zhang G, Luo M, Guo H, Zhan J, et al. Efficient multi-scale attention module with cross-spatial learning. In: ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE; 2023. p. 1–5. https://doi.org/10.1109/ICASSP49357.2023.10069516

32. Hu L, Li Y. Micro-YOLO: Exploring Efficient Methods to Compress CNN based Object Detection Model. In: ICAART (2). 2021. p. 151–8. https://doi.org/10.5220/0010234410151058

33. Li C, Li L, Jiang H, Weng K, Geng Y, Li L, et al. YOLOv6: A single-stage object detection framework for industrial applications. arXiv preprint arXiv:2209.02976. 2022.

34. Carion N, Massa F, Synnaeve G, Usunier N, Kirillov A, Zagoruyko S. End-to-End Object Detection with Transformers. In: European conference on computer vision. Cham: Springer International Publishing; 2020. p. 213–29. https://doi.org/10.1007/978-3-030-58452-8_13

35. Duan K, Bai S, Xie L, Qi H, Huang Q, Tian Q. Centernet: Keypoint triplets for object detection. In: Proc. IEEE/CVF Int. Conf. Computer Vision (ICCV). 2019. p. 6569–78. https://doi.org/10.1109/ICCV.2019.00667

FINANCING

None.

CONFLICT OF INTEREST

None.

AUTHORSHIP CONTRIBUTION

Conceptualization: Kamaleshwar T, Lokesh Kumar Paila, Sneja Godi, Christo Ananth, V.H. Abdullayev, R.G. Abaszade.

Data curation: Kamaleshwar T, Lokesh Kumar Paila, Sneja Godi, Christo Ananth, V.H. Abdullayev, R.G. Abaszade.

Formal analysis: Kamaleshwar T, Lokesh Kumar Paila, Sneja Godi, Christo Ananth, V.H. Abdullayev, R.G. Abaszade.

Drafting - original draft: Kamaleshwar T, Lokesh Kumar Paila, Sneja Godi, Christo Ananth, V.H. Abdullayev, R.G. Abaszade.

Writing - proofreading and editing: Kamaleshwar T, Lokesh Kumar Paila, Sneja Godi, Christo Ananth, V.H. Abdullayev, R.G. Abaszade.