Proc. of the European Safety and Reliability Conference (ESREL 2026)
Edited by José C. Matos, Paulo B. Lourenço, Michael Beer, Daniel Oliveira,
Edoardo Patelli, Hélder S. Sousa, Rui A. Silva and Monica Santamaria

doi: 10.3850/ESREL2026061419_esrel26-p27644-cd

Intelligent Video Surveillance Technology for Safety Risks in Oil and Gas Production Operations

Kun Tian

China University Of Petroleum(Beijing),China. Research Institute of Safety and Environment Technology, CNPC. E-mail: tiankuntk@cnpc.com.cn

Laibin Zhang

China University Of Petroleum(Beijing),China. E-mail: zhanglb@cup.edu.cn

Jinjiang Wang

China University Of Petroleum(Beijing),China. E-mail: jwang@cup.edu.cn

Shunyi Wang

Research Institute of Safety and Environment Technology, CNPC, China. E-mail: wangshunyi@ cnpc.com.cn

Zhuangyuan Hao

Research Institute of Safety and Environment Technology, CNPC, China. E-mail:haozhuangyuan@cnpc.com.cn

The petroleum and petrochemical industry involves a massive volume of high-risk operations. The traditional risk management model relies on on-site inspections by monitoring personnel, which suffers from numerous regulatory blind spots and incomplete coverage.As the temporal and spatial coverage of video surveillance continues to expand, massive amounts of image data are generated.In this study, 1.4 million images of high-risk operation sites were collected and compiled. Utilizing computer vision technology, a multi-mechanism dynamic identification method for safety hazards was constructed, integrating object detection, pose estimation, and object tracking. Specifically, the object detection model first outputs information such as target types (e.g., human bodies, cranes), keypoints (e.g., crane boom keypoints), and IDs. Spatiotemporal attention mechanisms were incorporated into the object detection model to enhance the recognition accuracy of multi-scale and small-scale targets within the images. Finally, based on the violation behaviors that need to be determined, algorithm reasoning logic was designed to enable the computer to distinguish and identify the behaviors of key targets from complex video images, automatically analyzing and extracting critical information from video sources.This study developed 40 types of video recognition algorithms for operational hazards, such as "personnel standing under a crane boom" and "working at heights without wearing a safety harness" (achieving an accuracy rate of ⩾90%). This enables the real-time, intelligent identification and analysis of violation behaviors, facilitating the transformation of operational risk management from a "human-based defense" to an integrated "human + technical + intelligent defense" model. It allows for early intervention in operational violations, thereby further elevating the standard of safety risk management.
Keywords: High-risk operations; Risk management; Computer vision; Video analysis; Hazard identification

1. Introduction

As a pillar industry of the national economy, the petroleum and petrochemical industry involves various high-risk scenarios in its production operations, such as hoisting, working at heights, and hot work. These operations are characterized by complex environments, concentrated hazards, and cumbersome procedures; once a safety accident occurs, it causes severe casualties, property damage, and ecological destruction.Currently, risk management for high-risk operations in the petroleum and petrochemical industry primarily relies on on-site inspections by monitoring personnel. Restricted by factors such as human stamina, professional expertise, and inspection scope, this traditional model faces prominent issues like numerous regulatory blind spots and incomplete coverage, making it difficult to meet the safety management demands of large-scale, high-intensity, and high-risk operations.

With the rapid development of video surveillance technology, high-definition cameras and infrared monitoring equipment have been widely deployed at high-risk operation sites in the petroleum and petrochemical industry. The continuous expansion of the temporal and spatial coverage of these monitoring systems has effectively addressed some of the shortcomings of traditional manual inspections.However, the explosive growth of massive image and video data from operation sites—where a single video channel can generate tens of thousands of images daily — has made manual processing highly inefficient. Furthermore, human fatigue and subjective judgment biases easily lead to missed or false detections of hazards, making it difficult to fully leverage the safety value of video surveillance data.Therefore, a critical issue urgently needing to be addressed in the safety management of high-risk operations within the petroleum and petrochemical industry is how to utilize advanced technologies for the intelligent analysis of massive operational image data to accurately and in real-time identify safety hazards during operations.

Deep learning technology has become a core supporting technology for risk prevention and control, hazard detection, and compliance monitoring in the industrial safety field, achieving significant progress in complex scene adaptation, detection accuracy, and the enhancement of intelligent levels (Wang et al., 2023; Hu et al., 2025).In the field of risk early warning for maintenance operations, researchers have constructed a dynamic risk assessment and hierarchical early warning framework based on the YOLOv5 model combined with Bayesian Networks (BN) and fuzzy set theory. This framework has successfully realized the real-time identification of risk factors in scenarios like compressor maintenance, effectively responding to changes in event probabilities and triggering hierarchical alarms (Wang et al., 2023).Addressing the pain points of hazard detection in oil and gas drilling, industrial production, and other scenarios, scholars have improved the YOLOv5 model by introducing technologies such as the SimAM attention mechanism and Adaptively Spatial Feature Fusion (ASFF). This has resolved issues like complex backgrounds and large variations in hazard scales, increasing the average recognition accuracy by 10.4%. Furthermore, they have developed smart wearable devices based on AR smart glasses, breaking through the regional limitations of traditional video surveillance and enhancing detection portability (Hu et al., 2025).Regarding personal protective equipment (PPE) compliance detection, researchers have employed YOLOv5 combined with the Openpose pose estimation algorithm and 1D Convolutional Neural Networks (1D-CNN) to achieve the automatic detection of behaviors such as standard safety helmet wearing and safety belt hook usage (Li et al., 2022). Meanwhile, a generic PPE compliance detection framework based on HigherHRNet pose estimation and classification models like MobileNetV2 can be adapted to detect protective equipment across 18 different body parts, providing flexible solutions for multi-industry scenarios (Vukicevic et al., 2022).

Hazard identification methods integrating multiple computer vision technologies have become a research hotspot. Although current studies have achieved breakthroughs in detection accuracy and environmental adaptability for specific scenarios, most existing methods target single operational scenarios or a limited number of hazard types. Moreover, they have not fully considered the particularities of high-risk operations in the petroleum and petrochemical industry, making it difficult to achieve the accurate and real-time identification of multiple types of hazards.Focusing on high-risk operational scenarios in the petroleum and petrochemical industry, this study constructs a multi-mechanism dynamic identification method for hazards that integrates multiple computer vision technologies. By developing hazard identification algorithms, this research achieves the real-time intelligent identification of and early intervention in operational violations. This promotes the transformation and upgrading of the industry's risk management model and provides a novel technical solution for ensuring the safety of high-risk operations in the petroleum and petrochemical sector.

2. Model Methodology and Improvement Schemes

2.1. Principles of the YOLOv8 Algorithm

The network architecture of YOLOv8 is divided into four core components (Glenn Jocher et al., 2023). The overall workflow is: Input Layer → Backbone → Neck → Head. The specific structure is illustrated in Figure 1.

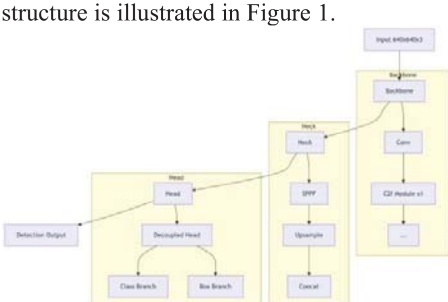

Fig. 1 Network architecture of YOLOv8

2.2. Algorithm Improvement Strategies

2.2.1. Introducing Attention Mechanisms to Suppress Background Noise

YOLOv8 utilizes PAN-FPN for multi-scale feature fusion; however, the traditional concatenation (concat) fusion method fails to distinguish the importance of features across different scales, making the low-level detail features of small targets easily overwhelmed by high-level semantic features.By inserting attention modules at the feature fusion nodes in the Neck section to allocate weights to features from different scales and branches — thereby reinforcing useful features and suppressing redundant ones — the model can automatically focus on the key features of small targets amid background noise, thus enhancing fusion efficiency.

A Convolutional Block Attention Module (CBAM), representing a dual channel-spatial attention mechanism, is employed in the enhancement stage following feature fusion. Its channel attention branch performs global pooling operations on each channel of the feature map, leveraging a Multi-Layer Perceptron (MLP) to learn channel weights to reinforce critical channels (e.g., texture channels of small targets). Conversely, the spatial attention branch applies pooling treatments to each position of the feature map, utilizing convolutions to learn spatial weights to enhance the feature representation of the target region.

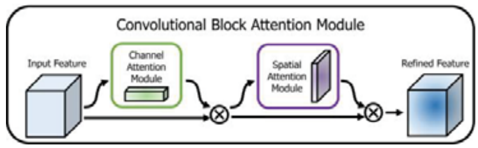

Fig. 2: Attention mechanism module

The improvement advantages of the CBAM attention module are primarily manifested in two aspects. First, it accurately learns and elevates the channel weights of low-level detail features (such as textures of small targets) through the channel attention branch, while simultaneously reinforcing the feature representation of small target regions combined with the spatial attention branch. This effectively resolves the issue within the PAN-FPN structure where low-level features are easily "covered" by high-level semantic features, significantly enhancing the feature discriminability of small targets.Second, the module can precisely focus on target regions via spatial attention and reinforce key feature channels via channel attention. Even in complex scenarios involving occlusion or blurring, it can improve the distinction between target features and the background, demonstrating excellent detection accuracy in practical application scenarios involving small targets (e.g., smoke/fire, gas detectors) and equipment/facility occlusions typical in special operation video surveillance within the petroleum and petrochemical industry.

2.2.2. Replacing the Loss Function with WIoU for Bounding Boxes to Improve Regression Accuracy

YOLOv8 employs CIoU Loss for bounding box regression; however, CIoU suffers from issues such as an unreasonable aspect ratio penalty term and a slow convergence speed.

WIoU (Weighted IoU), by introducing dynamic weights, prioritizes the model's optimization of high-quality anchor boxes, thereby improving regression accuracy.

The core of WIoU lies in adding a dynamic weighting factor to the IoU loss. Based on the IoU magnitude between the predicted box and the ground truth box, it assigns larger weights to high-quality predicted boxes (those with large IoU), allowing the model to focus on optimizing samples that are closer to the ground truth. The calculation method for WIoUv2 is as follows:

$$L_{WIoU} = 1 - IoU + \frac{p^2(b, b^{gt})}{c^2} + \alpha$$

where the weighting factor α is dynamically determined by the IoU: $\alpha = (1 - IoU)^{\gamma}$ (γ is a hyperparameter), which addresses the flaw of a fixed α in CIoU.

The advantages of WIoU over CIoU are mainly reflected in three aspects. First, superior convergence speed; by increasing the weight allocation for high-quality samples, this method enables the model to rapidly fit the ground truth boxes in the early stages of training, significantly shortening the training convergence cycle.Second, significantly improved regression accuracy; it exhibits greater positioning accuracy in target localization tasks such as human keypoint and crane keypoint detection. Third, further enhanced robustness; it yields better fitting results for abnormal samples like extremely small or extremely large targets, effectively reducing regression errors in extreme scenarios and ensuring the model's stable performance across complex data distributions.

3. Video Recognition Algorithm for Safety Hazards

Hoisting operations represent a typical high-risk task in the petroleum and petrochemical industry, directly impacting the life safety of operators. During hoisting operations, affected by the movement of the crane boom and hook, safety accidents such as rope decoupling, unhooking, and rope breakage are highly prone to occur. Among these, personnel illegally entering hazardous zones is the primary cause of casualties, mainly involving two typical scenarios: first, personnel located within the rotation radius of the crane boom (hereafter referred to as "personnel standing under the crane boom"), who may suffer from the boom's slewing impact or falling lifted objects; second, personnel located directly beneath the lifted object (hereafter referred to as "personnel standing under the hook"), directly facing the risk of the object unhooking and falling. Because their risk mechanisms differ, identification algorithms must be established separately. This study developed corresponding identification algorithms for these two types of hazards (see items 25 and 29 in Table 8), and the following details the algorithm reasoning logic taking "personnel standing under the crane boom" as an example.

Currently, a large number of surveillance cameras have been deployed at operation sites, accumulating rich video and image data, which provides data support for determining the behavior of personnel standing under a crane boom using vision technology. Existing technologies can detect hooks and human bodies through general object detection network models, but relying solely on conventional object detection methods suffers from insufficient detection accuracy. By introducing an attention module into the detection network and optimizing the loss function, this study significantly improved recognition accuracy. Concurrently, deep learning methods are employed to directly regress spatial coordinates, clarifying the spatial positional relationship between targets. Through the conversion from 3D to 2D coordinates, an end-to-end neural network method directly outputs spatial positional relationships, achieving real-time detection and timely warning for the behavior of "personnel standing under the crane boom," thereby facilitating the rapid evacuation of personnel below and reducing the likelihood of accidents.

3.1. Data Collection and Preprocessing

Staged videos (including core targets like cranes and booms) were collected from oilfield sites. Corel Video Studio Pro X5 video editing software was used to extract clips containing violation behaviors. The OpenCV image processing library was utilized to split the useful video clips into image frames, and highly repetitive frames were removed via programming to reduce data redundancy. Subsequently, data augmentation was performed through random flipping, scaling, and Mosaic operations to enhance the model's generalization ability. Finally, the LabelImg annotation tool was used to complete the visual data annotation. Five keypoints were used for structural annotation of the crane's core skeleton, with specific physical mapping definitions as follows: Keypoint 1 represents the root of the boom, Keypoint 2 represents the crane's slewing center (i.e., the vertical projection of Keypoint 1 on the chassis), Keypoint 3 represents the tip of the boom, Keypoint 4 represents the hook and hoisted object, and Keypoint 5 represents the vertical projection of the boom tip on the ground. The annotation diagram is shown in Figure 3.

Fig. 3 Annotation Schematic Diagram

This study constructed an original image database of oil and gas production operations containing 1.4 million images. To efficiently utilize this massive data and ensure algorithm accuracy, a hierarchical processing strategy was adopted. First, 125,000 keyframe images were extracted from the original database for accurate manual annotation. For annotation quality evaluation, a quality control mechanism of 'double-blind annotation + expert review' was established. Each image was independently annotated by two annotators; if their Intersection over Union (IoU) for the same target's bounding box was > 0.85 and the categories matched, it was directly entered into the database. Disputed samples (accounting for approximately 14.2%) were handed over to senior safety experts for cross-review and arbitration. Ultimately, the inter-rater agreement of the dataset reached 96.5%, ensuring high confidence.

For the 1.275 million unannotated data, a semi-supervised learning strategy based on a Teacher-Student architecture was employed. The initial model trained on the finely annotated dataset was used to generate pseudo-labels. After strict confidence filtering, these pseudo-labeled data were mixed with the finely annotated data for iterative training, significantly improving the model's generalization performance under unseen lighting conditions and complex industrial backgrounds.

Targeting the 40 developed operational hazard recognition algorithms, the annotated dataset covers five major categories: general scenarios, working at heights, hot work, hoisting operations, and confined space operations. The distribution statistics of the overall categorical annotated data are shown in Table 1, which includes the special hoisting experimental data emphasized in Section 3.3.

Table 1: Distribution Statistics of the Full-Category Hazard Annotated Dataset

No.	**Major Hazard Category**	**Number of Algorithms Included**	**Number of Annotated Samples (Images)**
1	General Scenarios	15	45,200
2	Working at Heights	2	18,500
3	Hot Work	6	21,300
4	Hoisting Operations	14	25,600
5	Confined Spaces	3	14,400
Total		40	125,000

Ultimately, 5,834 images related to personnel standing under the boom were collected, completing the annotations for crane bounding boxes, human bounding boxes, and crane keypoints. Meanwhile, 27,823 construction background images from oil and gas field enterprises were selected as negative samples to address the issue of class distribution imbalance during model training and to improve scene robustness. All data were divided into a training set and a testing set at an 8:2 ratio. The annotation labels and their corresponding quantities are shown in Table 2.

Table 2: Labels and Quantities

Serial Number	Label	Quantity
1	Crane Rectangular Box	5064
2	Human Rectangular Box	12621
3	Crane Key Points	25320

3.2. Experimental Environment and Parameters

In this experiment, model training was performed on an NVIDIA RTX 4080 Ti graphics card. PyCharm was used as the primary development tool, and the model construction, training, and validation were implemented based on the PyTorch deep learning framework. The training process utilized the Adam optimizer with 100 epochs and a batch size of 10. The initial learning rate was set to 0.001, which was adjusted to 0.0005 after the 10th epoch. The IoU threshold was set to 0.5. The experimental dataset was divided into a training set and a testing set at a ratio of 8:2, and the image resolution was uniformly resized to 640×640.

3.3. Algorithm Model Reasoning Logic

The RGB data of the image is fed into the improved YOLOv8 detection model, which outputs the prediction results of the crane bounding box, human bounding box, and the detection results of the 5 keypoints of the crane.

Fig. 4 Detection Result Schematic Diagram

For personnel standing under the crane boom, the specific reasoning logic is as follows:

1. Operational status determination: Determine whether a hoisting operation is underway based on the boom's deployed state. If the crane detection confidence meets the preset threshold, and the intersection angle between the line segment connecting crane keypoints 1-3 and the line segment connecting keypoints 2-5 is greater than 30°, it is determined that a hoisting operation is currently taking place.

2. Hazard spatial determination: Determine if there is a hazard of "personnel standing under the crane boom." Meeting either of the following two conditions is recorded as one violation state:

① Vertical projection area verification: First, extract the personnel detection box to determine if the person is in a valid working posture (the aspect ratio is between 0.2 and 0.6 to accommodate both standing and squatting postures). Second, taking the bottom center point of this human box as the center and 0.75 times the width of the personnel box as the diameter to draw a circle, if this circle spatially intersects the line segment formed by crane keypoints 2 and 5, a determination is triggered.

② Rotating sector area verification: Similarly, extract the human box with a valid posture. If the bottom center point of this human box is located within the sector risk zone of the boom's mechanical rotation, a determination is triggered. (Calculation method of the sector risk zone: using crane keypoint 2 as the center, the length of the line segment from keypoint 2 to 5 as the radius, and the line connecting 2 and 5 as the central axis, expanded by 20° to each side to form a sector area) .

3. Dynamic alert triggering mechanism: To filter occasional detection jitter, the system introduces a cumulative confidence mechanism. The initial violation confidence is set to 0 (the minimum lower limit is 0). During continuous frame detection, if a violation state is identified in a single frame, the confidence increases by 0.2; if not identified, it decreases by 0.1. When the cumulative violation confidence exceeds 1.0, the system immediately and automatically triggers an on-site alert.

3.3.1. Coping Strategies for Complex Scenarios

Addressing challenges such as overlapping or occluded operators and strong perspective distortion caused by surveillance angles at high-risk operation sites, this algorithm adopted targeted optimization measures. Regarding target overlapping, by introducing the ByteTrack algorithm, a unique ID is assigned to each detected human target. Leveraging the spatiotemporal continuity features of the targets, the system ensures that it can continuously and accurately track and determine the movement trajectory of the bottom center point of each human body even after brief periods of intersection or occlusion, effectively reducing the risk of recognition interruption or ID switching caused by target overlap.

Aiming at the planar assumption limitations of traditional Homography Transformation in rugged and uneven terrains such as onshore desert well sites and mud pits, engineering optimizations were implemented during the spatial calibration phase in this study. Building on the initial homography matrix mapping using on-site reference objects (e.g., outrigger spans), Local Ground Plane Approximation and a human prior scale compensation mechanism were introduced. Specifically, the relative height change of operators within the detection box is extracted as a local depth reference to dynamically correct physical coordinate mapping deviations caused by ground undulations. For computational residuals potentially brought by extreme terrain undulations, the system employs a multi-frame spatiotemporal trajectory smoothing mechanism for fault-tolerant filtering, significantly enhancing the robustness of the hazard determination logic in complex industrial scenarios.

3.4. Evaluation Metrics

Based on single video stream access, algorithm performance is comprehensively evaluated using three metrics: Precision, Recall, and model inference speed (FPS). The validation method involves randomly extracting several video segments on-site for testing. The calculation methods for each metric are shown in Table 3.

Table 3: Metric Calculation Methods

Serial Number	Evaluation Indicator	Calculation Method
1	Accuracy (Prec)	Accuracy = Number of Correct Cases / (Number of Correct Cases + Number of Incorrect Cases)
2	Recall Rate (Recall)	Recall Rate = (Total Number of Events - Number of Missed Events) / Total Number of Events
3	Inference Speed (FPS)	Number of images inferred per second, feedback from the background

3.5. Comparative Experiments

Taking an Intersection over Union (IoU) of 0.50 as the determination threshold for successful detection, Precision and Recall are used as the core accuracy evaluation metrics to measure the comprehensive performance of the trained model across all categories. Faster-RCNN, original YOLOv8, and improved YOLOv8 were selected for comparative experiments. The experimental results are shown in Table 4.

Table 4: Comparison of Experimental Results

Model	Precision	Recall Rate	Inference Speed
Faster-RCNN	85.7%	83.2%	Slow
YOLOv8	87.8%	86.3%	Fast
Improved YOLOv8	**92.4%**	**90.6%**	**Fast (Real-time, comparable to the original)**

The improved YOLOv8 verified the effectiveness of the CBAM attention mechanism and the WIoU loss function. The CBAM module strengthened the feature distinction between small targets and complex backgrounds, while WIoU optimized the bounding box regression accuracy. Their synergy resolved the detection pain points of large target scale variations and complex backgrounds in petroleum and petrochemical scenarios.

3.5.1. Ablation Experiments and Failure Case Analysis

To further verify the genuine contributions of the improved modules and eliminate random influences during the training process, this study introduced a significance testing mechanism in the ablation experiments. Each scheme underwent 5 independent repetitive trainings under identical experimental hyperparameters. The average values of Precision and Recall were taken as the final results for comparison (see Table 5).

Table 5: Results of Ablation Experiments on Algorithm Modules

Scheme	Improvement	Precision	Recall
Baseline	Original YOLOv8	87.8%	86.3%
Exp 1	YOLOv8 + CBAM	90.1%	88.5%
Exp 2	YOLOv8 + WIoU	89.4%	87.9%
Proposed	YOLOv8 + CBAM + WIoU	92.4%	90.6%

Meanwhile, a paired-sample t-test was conducted on the 5 independent Precision results of the Baseline (original YOLOv8) and Proposed (improved model) schemes. Statistical analysis results indicated that the accuracy improvement of the modified model carries highly significant statistical meaning ($p < 0.01$). This not only excludes random effects at the statistical level but also fully proves the stable gains of the combination of CBAM and WIoU in enhancing small target detection capability against complex backgrounds.

Failure case explanation: In situations of extreme strong light shining directly into the camera or crane keypoints being completely occluded by large equipment (occlusion rate > 80%), the model may experience missed detections or keypoint drifting. This points to the future research direction of introducing multi-view stereo vision monitoring.

3.5.2. Dataset Scale Ablation Experiments

To verify the contribution of the 1.4 million original images collected in this study and the 125,000 finely selected and annotated data to model performance, an ablation experiment on dataset scale was conducted. By proportionally extracting 20% (25,000 images), 50% (62,500 images), and 100% of the data from the 125,000 finely annotated dataset for training, the model's performance on the same testing set was observed. The experimental results are shown in Table 6.

Table 6: Results of Dataset Scale Ablation Experiments

Experimental Scheme	Proportion of Annotated Data	Annotated Sample Size (Images)	Precision	Recall	mAP @ 0.5
Data-Exp 1	20%	25,000	81.4 %	78.5%	79.2 %
Data-Exp 2	50%	62,500	88.2 %	85.6%	86.8 %
Data-Exp 3	100%	125,000	92.4 %	90.6%	91.5 %

Experimental results showed that as the annotated data scale increased from 25,000 to 125,000 images, the model's Precision improved by 11 percentage points, and Recall improved by 12.1 percentage points. This indicates that in complex environments with high background noise and variable target scales, such as petroleum and petrochemical operation sites, relying on small sample sizes makes it difficult to cover the feature distributions required for the 40 hazard algorithms. Furthermore, by introducing 1.275 million unannotated images for semi-supervised pre-training, the initial performance of the model at the Data-Exp 1 stage was already superior to models of the same scale without pre-training. This experiment fully demonstrates the core value of building large-scale industry datasets for resolving the "long tail effect" in hazard detection and enhancing the reliability of safety-critical systems.

3.5.3. Statistical Rigor Analysis

To eliminate sporadic impacts, the improved model underwent 5 independent training runs. Results showed a mean Precision of 92.4% with a standard deviation of $\sigma = 0.35\%$, exhibiting extremely high training stability.

IoU Threshold Sensitivity: The experiment tested the impact of different IoU thresholds on mAP. When the IoU threshold fluctuated within the range of [0.5, 0.75], the Precision fluctuation range remained within 2.1%, proving the robustness of the WIoU loss function in bounding box regression.

3.5.4. Environmental Adaptability and Generalization Validation

By extracting specific scenario samples from the validation set, the model's generalization performance under diverse sites and complex meteorological conditions was comparatively analyzed. Results indicated that the precision fluctuation between onshore deserts (92.6%) and offshore platforms (91.8%) was only 0.8%. Even in foggy weather with low visibility (90.5%), the model still exhibited exceptionally strong robustness. This further confirms the stability of the improved algorithm in cross-regional and cross-temporal applications. The quantitative validation results are shown in Table 7.

Table 7: Model Generalization Validation Across Different Working Environments and Time Periods

Validation Dimension	**Scenario Classification**	**Sample Size (Images)**	**Precision (Prec)**
Geographical Environment	Onshore desert well site	3,200	92.6%
	Offshore drilling platform	2,800	91.8%
Lighting/Weather	Dawn/Dusk (Low light)	2,500	91.2%
	Fog/Dust (Low visibility)	1,800	90.5%
	Sunny (Strong light)	3,700	92.8%

4. Field Application of Intelligent Hazard Recognition Algorithms

Based on the aforementioned research methods, 1.4 million images from enterprise sites were collected, and 40 types of operational hazard recognition algorithms were successfully constructed, covering five major core scenarios: general scenarios, working at heights, hot work, hoisting operations, and confined space operations. The specific algorithm list is shown in Table 8.

Table 8 List of Hazard Identification Algorithms

Serial Number	**Scenario**	**Algorithm**
1	General	Helmet Wearing
2		Safe Dressing
3		Open Flame
4		Smoke
5		Smoking
6		Person Falling to the Ground
7		Area Intrusion
8		Absence from Post
9		Sleeping on Duty
10		Answering/Making Phone Calls
11		Oil Leakage
12		Monitoring Screen Occlusion Recognition
13		Fire/Safety Exit Blockage
14		Heavy Vehicle Intrusion
15		Personnel Gathering
16	High-Altitude Operations	Not Wearing Safety Belt in High-Altitude Operations
17		High-Altitude Operations Without Supervision
18	Hot Work	No Fire-Fighting Facilities on Site
19		Flammable Materials Near the Operation Site
20		Exceeding the Limit of Personnel in the Operation Area
21		Oxygen Cylinder and Acetylene Cylinder Too Close
22		Personnel Not Wearing Protective Masks
23		Acetylene Cylinder Laid Horizontally
24	Hoisting Operations	No Warning Line Set for Hoisting Operations
25		Personnel Standing Under the Hook of Girder Bridge Crane
26		Hoisting in Bad Weather
27		No Warning Tape for Hoisting
28		Hoisting Operations Without Commanders
29		Personnel Standing Under the Crane Arm
30		Driver Leaving During Operation
31		Crane Arm Not Retracted After Operation
32		Crane Machinery Too Close to Transmission Lines
33		Outriggers Suspended
34		Outriggers Sinking
35		Outriggers Near Cable Trench/Pipe Trench Manhole Covers
36		Outriggers Not Extended

37		Outriggers Without Base Plates
38	Confined Space Operations	Holding Lifted Objects by Hand
39		Not Wearing Portable Gas Alarm
40		Dual Supervision at Confined Space Entrance

In this study, the constructed intelligent video recognition algorithm models were encapsulated into an algorithm model repository and deployed on hardware carriers such as video analysis computing boxes and video analysis servers that provide video services and AI intelligent analysis services. These hardware devices can be flexibly deployed at central nodes, edge nodes, or front-end locations to adapt to the deployment requirements of different operation sites.

The core workflow for field application is as follows: 1) Data collection: On-site cameras capture operational video data in real time and upload it to the algorithm model repository; 2) Algorithm configuration: The server side connects with the streaming media platform or cameras, configures corresponding algorithm models for each camera channel, and intelligently analyzes the video of that channel by calling the models in the computing-end algorithm model repository via configuration; 3) Early warning and closed-loop management: After model analysis, early warning results are generated and returned to the server for display. The server prompts relevant personnel to handle the warning events through various means, achieving closed-loop management of the events.

4.1. Alarm Closed-Loop and Ethical Considerations

To balance the efficiency of oil and gas production operations with the intensity of safety supervision, this system has established a three-tier "warning-verification-intervention" closed-loop management mechanism in practical applications. In terms of building system robustness and trustworthiness, by introducing a cumulative confidence fault-tolerant mechanism, the system effectively filters out false alarms caused by transient factors such as passing birds or fluctuating environmental lighting. The measured false alarm rate is stably controlled within 3%, significantly enhancing the technical trust of grassroots monitoring personnel in the intelligent supervision system.

In the design of the manual intervention process, the system takes into account both response speed and the principle of hierarchical disposal: once the AI algorithm identifies a safety hazard, the backend voice intercom system will automatically issue a warning tone to the operation site within 2 seconds to achieve an instant reminder. If the violation state persists for more than 10 seconds without correction, the early warning information will be automatically and synchronously pushed to the safety director and the management end for intervention. This collaborative mode of "intelligent defense as the mainstay, human defense as coordination" avoids unnecessary facial information collection through end-to-end recognition in its technical implementation. This not only protects operators' privacy at the ethical level but also ensures the efficient resolution and closed-loop management of safety risks in high-risk operations.

In actual applications, the system allows safety management personnel to dynamically adjust the alarm trigger threshold (currently defaulted to confidence > 1.0) based on the risk level of the operational scenarios. For high-risk operational areas (such as hoisting operations), the trigger threshold can be appropriately lowered to increase sensitivity; for general operational areas, the threshold can be appropriately raised to reduce interference. This flexible configuration mechanism has been verified in on-site deployments, allowing users to adjust the cumulative confidence threshold within the range of 0.8-1.5 according to actual needs, thereby balancing the intensity of safety supervision with operational efficiency.

5. Conclusion

Addressing the issues of numerous blind spots, low efficiency, and the difficulty in effectively mining and utilizing massive video data in the traditional manual supervision of high-risk operations in the petroleum and petrochemical industry , this study, based on the YOLOv8 algorithm, constructed an improved detection model adapted to industry scenarios. This was achieved by introducing the CBAM channel-spatial dual attention module in the Neck section to enhance the feature distinction between small targets and backgrounds, and by replacing the regression loss function with WIoU to improve bounding box positioning accuracy and convergence speed.Experimental results demonstrate that the comprehensive performance of the improved YOLOv8 model is superior to that of Faster-RCNN and the original YOLOv8. Its precision reaches 92.4%, its recall reaches 90.6%, and its inference speed meets the requirements for real-time monitoring (>30 FPS).

Focusing on core operational scenarios in oil and gas fields such as hoisting, working at heights, and hot work, this study proposes a multi-mechanism dynamic identification scheme for safety hazards that integrates object detection, pose estimation, and object tracking. Furthermore, it developed 40 types of operational hazard recognition algorithms—such as personnel standing under the crane boom, personnel standing under the hook, and working at heights without wearing a safety harness (achieving an accuracy rate of ⩾90%). Through a cloud-edge-end collaborative architecture, these algorithms are encapsulated and deployed on hardware carriers such as video analysis computing boxes and servers, achieving the real-time recognition, warning, and closed-loop management of violation behaviors.This research effectively promotes the transformation and upgrading of risk management in the petroleum and petrochemical industry from a "human-based defense" to an integrated "human + technical + intelligent defense" model. It provides an efficient and precise technical solution for ensuring the safety of high-risk operations, and possesses significant industrial application prospects and practical value.

Although this study has alleviated the impacts of perspective distortion and target overlap to a certain extent through homography transformation and object tracking algorithms, the monocular vision scheme still faces a physical bottleneck of fluctuating recognition accuracy when dealing with extremely complex occlusions (e.g., occlusion rate > 80%) or highly challenging strong perspective angles due to the lack of depth information.In future research, the team plans to introduce multi-camera feature fusion technology based on multi-view stereo vision. By establishing a digital twin space of the operation site through 3D calibration, and utilizing multi-dimensional spatial projection and 3D coordinate reconstruction, we aim to further eliminate visual blind spots and resolve the issue of target loss in highly overlapping scenarios. This will continuously enhance the reliability and rigor of safety-critical systems in extreme industrial environments.

Acknowledgments

This work was supported by the following research projects:

- China National Petroleum Corporation (CNPC) Scientific Research Projects (Nos. 2023DJ6508,2023DJ7112);

References

Chian, E.Y.T., Goh, Y.M., Tian, J. and Guo, B.H.W. (2022). Dynamic identification of crane load fall zone: A computer vision approach. Safety Science, 156, pp.105904.

Hu, J., Zhang, L., Hu, Y. and Li, Z. (2025). YOLOv5s-SimAM-ASFF-based oil and gas drilling and production wellsite safety hazard identification and smart wearable device development. Science Technology and Engineering, 25(16), pp.7004–7012.

Ji, Z., Zhou, Y., Zhang, Y. and Shi, K. (2024). Industrial site unsafe behavior detection based on improved YOLOv5. China Safety Science Journal, 34(7), pp.38–43.

Li, J., Zhao, X., Zhou, G. and Zhang, M. (2022). Standardized use inspection of workers' personal protective equipment based on deep learning. Safety Science, 150, pp.105689.

Vukicevic, A.M., Djapan, M., Isailovic, V. and Milasinovic, D. (2022). Generic compliance of industrial PPE by using deep learning techniques. Safety Science, 148, pp.105646.

Glenn Jocher, Ayush Chaurasia, Jing Qiu. Ultralytics YOLOv8[CP/OL]. V8.0.0. 2023. https://github.com/ultralytics/ultralytics.

Wang, J., Guan, P., Chen, Z. and Ge, W. (2023). Intelligent warning of risk during maintenance operations based on deep learning. China Safety Science Journal, 33(10), pp.16–22.

Chen, H., Hou, L., Zhang, G.K. and Wu, S. (2023). Using Context-Guided data Augmentation, lightweight CNN, and proximity detection techniques to improve site safety monitoring under occlusion conditions. Safety Science, 158, pp.105958.

Duan, P.S., Zhou, J.L. (2022). Description Method of Construction Workers' Unsafe Behaviors Based on Posture Characteristics. Safety and Environmental Engineering, 29(3), pp.1-8.

Liu, Y., Jiao, S.J. (2022). Application of ST-GCN in unsafe action identification of construction workers. China Safety Science Journal, 32(4), pp.30-35.