Design of Monitoring System for Coke Quenching Lifting Machine Based on Deep Learning

Xinzhong Yu, Shaochuan Xu, Ting Wang

Abstract—**In an effort to evolve dry quenching coke can hoists from conventional monitoring practices to a more sophisticated intelligent dynamic monitoring system, this paper presents a novel approach based on deep learning methodologies. Initially, it introduces an advanced instance segmentation framework, You Only Look Once version 11 Nano with compact local instance segmentation, which represents an enhancement of the existing You Only Look Once version 11 Nano with compact local instance segmentation framework. You Only Look Once version 11 Nano with compact local instance segmentation achieves a lighter model footprint through modifications in the backbone network and feature fusion techniques, further reinforced by the integration of a local feature attention mechanism to enhance its ability to discern local features. Subsequently, utilizing the improved model as the foundation for instance segmentation, this paper defines eight distinct operational states of the hoist, informed by the detailed insights generated by the model. The research culminates in the design of a comprehensive dry quenching monitoring system, achieved by synthesizing multi-source information to establish state discrimination criteria, complemented by the incorporation of temporal analysis. Experimental findings underscore the system's exemplary performance in terms of accuracy, real-time efficiency, and lightweight design, demonstrating its capability to precisely and promptly ascertain the hoist's operational status, thereby providing robust technical support for the safe production processes in dry quenching.**

Index Terms—**Dry Quenching Coke Can Hoist, Deep Learning, Multi-Source Data Fusion, Intelligent Dynamic Monitoring System.**

I. Introduction

THE dry quenching hoist plays a pivotal role in the dry quenching process, responsible for lifting fully loaded red coke cans from the ground to the top of the dry quenching furnace and ensuring precise unloading [1]. The operational stability and reliability of this hoist are paramount, as they directly influence the overall efficiency, product quality, and safety of the dry quenching system. Consequently, effectively monitoring the hoist's condition and rapidly responding to any anomalies is crucial for optimizing the efficiency of the entire dry quenching process.

The conventional approach to monitoring dry quenching coke can hoists has relied on manual inspection and basic sensors. However, this method is constrained by the skill level, experience, and subjective judgment of the operators, making it challenging to provide real-time, comprehensive monitoring and prone to missing hidden faults. Although simple sensors like temperature and vibration sensors can offer real-time data on certain parameters, they provide only a one-dimensional view, which is not enough for a thorough and precise evaluation of the hoist's operational state. With advancements in computer technology, deep learning, a significant subset of AI, has achieved remarkable progress in areas such as target detection, semantic segmentation, and instance segmentation[2]-[4]. Its ability to learn features and recognize patterns provides an innovative solution for hoist condition monitoring.

The field of object detection aims to pinpoint and categorize specific objects within images or videos. Deep learning-based object detection emerged in 2014 with the introduction of R-CNN by Girshick and colleagues, marking the first use of CNNs in this context. The method involved using Selective Search to propose regions, AlexNet to extract features, and SVM for classification[5]. YOLO, introduced by Redmon and team in 2016, was a game-changer, treating detection as a regression issue and enabling real-time detection[6]. In 2020, Facebook developed the first completely end-to-end detection model, leveraging Transformer to understand global contexts and eliminating the need for NMS post-processing[7]. YOLOv10, presented by Wang and co-authors in 2024, further advanced the field by achieving end-to-end, real-time detection[8]. Object detection has since become a mature and widely applied technology across various domains. Notably, in 2025, Xiaoliang Zhu and his team's LPST-GCN model excelled in infrared action recognition, while Haiying Qi and others' IC-YOLOv5 algorithm demonstrated high accuracy in vehicle recognition under complex scenarios[9]-[10].

Semantic segmentation is a key task in computer vision, where each pixel in an image is categorized. The field has seen a major transition from traditional approaches to deep learning, evolving towards more efficient, accurate, and end-to-end methods. In 2015, the U-Net model, introduced by Ronneberger and colleagues, set a standard in medical image segmentation with its innovative encoding-decoding and skip connection architecture[11]. In 2017, Zhao et al. introduced PSPNet, utilizing a Pyramid Pooling Module (PPM) to effectively integrate global contextual information[12]. In 2023, SegmentAnything, which was trained on a dataset containing a billion items and functions as a promptable base model, advanced semantic segmentation into a new, more mature stage [13]. Semantic segmentation has become a well-established technology across various domains. A case in point is the I-U-Net model introduced by Yun Bai in 2025, which demonstrated remarkable precision in identifying road cracks[14].

Instance segmentation involves identifying not just the

Manuscript received May 2, 2025; revised July 18, 2025.

Xinzhong Yu is a postgraduate student of the University of Science and Technology Liaoning, Anshan, 114051, China(e-mail:1953496819@qq.com).

Shaochuan Xu is a Professor at the University of Science and Technology Liaoning, Anshan, 114051, China(Corresponding author phone: 86-0412-5929747; e-mail: shaochuanxu@163.com).

Ting Wang is a postgraduate student of Guangxi Normal University, Guilin, 541006, China(e-mail:2239729472@qq.com).

(a) Abnormal Hook Opening Malfunction (b) Premature Hook Grabbing Malfunction (c) Abnormal Hook Grabbing Malfunction

Fig. 1: Hoist Malfunction Diagram

type of object in an image, but also distinguishing between separate instances of the same object. The field is moving towards creating systems that work from start to finish in real-time. A key milestone was set in 2017 when He and his team built upon the Faster R-CNN framework to create Mask R-CNN, which seamlessly combines object detection with instance segmentation[15]. In the year 2020, a significant advancement in instance segmentation was made by Wang et al. with the introduction of the SOLO framework. This framework employs an innovative "instance category grid" to facilitate anchor-free, end-to-end instance segmentation, marking a notable development in the field [16]. In 2022, the Mask2Former model was introduced by Bowen Cheng and his team, utilizing masked attention to effectively perform panoptic segmentation[17]. Instance segmentation finds applications across various fields and has reached a relatively mature stage of development. A case in point is the EIS-YOLO model introduced by Xu Tan in 2025, which demonstrates remarkable precision in identifying surface damage on vehicles[18].

In the context of dry quenching coke can elevator operations, the nuanced changes in the elevator's state are not effectively detectable through conventional target detection technologies, which are limited to approximate object localization and categorization. Although semantic segmentation provides pixel-level category determination, it lacks the capability to differentiate between distinct instances within the same category. Instance segmentation, however, offers a comprehensive approach by incorporating both semantic classification and individual instance recognition. This enables precise identification of the elevator's hanging ears, a critical factor in assessing the grab's condition. Consequently, instance segmentation emerges as the optimal technique for monitoring the state of dry quenching coke can elevators.

Despite advancements in model simplification and performance enhancement within the field of instance segmentation, practical applications continue to encounter limitations. The objective of this study is to facilitate accurate detection of the coke can elevator's operational status, aligning with specific practical requirements. The contributions of this study are as follows:

The research introduces a novel instance segmentation framework, YOLOv11n-CL-seg, which builds upon the YOLOv11n-seg architecture. It achieves a more streamlined model through the substitution of the backbone network and feature fusion techniques. Furthermore, the incorporation of a local feature attention mechanism enhances the model's capacity to effectively capture and concentrate on local feature details. This optimization strategy has been demonstrated to significantly improve the model's focus and analytical precision regarding the intricate aspects of elevator steel structures, thereby substantially enhancing overall detection performance.

The research defines eight distinct operational states of the dry quenching coke can elevator, encompassing the entire operational workflow. Utilizing the mask map characteristics derived from instance segmentation, the study integrates multiple data sources, including the quantity and location of the ears and encoder parameters, to establish a robust framework for state discrimination. Furthermore, incorporating a temporal dimension allows for the analysis of state duration and transition logic, culminating in the design of a comprehensive dry quenching monitoring system. This system enables real-time, precise determination of the elevator's operational state, thereby providing significant technical support for the safe production processes in dry quenching operations.

II. Related Work

A. Dry Quenching Coke Drum Work Process

In the operational workflow of dry quenching, the coke can elevator initially deposits an empty can onto the production line. It subsequently retrieves a can filled with red coke and transports it to a predetermined location. Upon the precise positioning of the filled can by the conveyor system, the elevator's grab mechanism gradually lowers and securely attaches to the can's ear. Following a successful grab, the elevator elevates the can for further processing. Throughout this procedure, it is imperative to continuously monitor the status of the dry quenching coke can elevator to ensure operational safety and efficiency. Particular attention is paid to the successful release and acquisition of the grab hook, as any detection of failure in these actions necessitates an immediate alarm to prompt rapid response from the operational personnel.

B. Analysis of Hoist Malfunction

In the practical application of dry coke quenching processes, the hoist machinery is susceptible to three predominant failure modes. The first category of failure involves an irregular descent of the hook during the empty coke drum release, resulting in an inability to effectively disengage from the pendant, as depicted in Figure1a. Due

to the identical visual characteristics with the normal unseparated state, this malfunction cannot be discerned through static imagery or individual video frames. Accurate identification necessitates the application of time-series analysis in conjunction with established thresholds for action duration. The second category of failure is characterized by the hook initiating the grasping procedure prematurely, before the pendant has reached the targeted positioning zone. Referencing Figure1b, the system is equipped to rapidly diagnose issues by establishing a predefined region of interest (ROI) for the pendant, ensuring real-time precision in its placement. The third category of failure involves the hook's misalignment, resulting in it grasping only the outer portion of the pendant, as depicted in Figure1c. For the intelligent diagnosis of such failures, a multidimensional approach is essential. This entails not only analyzing the relative positioning of the hook and pendant within individual frames but also considering their motion trajectories within the broader context of the operational process. The accurate identification of these three failure types is paramount to ensuring the safety and stability of the dry quenching coke production process.

C. Difficulty Analysis in Detection

This article centers on the dry quenching coke can elevator and its integral hanging ears. The system is designed for the real-time monitoring of the elevator's status. Upon detection of a fault, an alarm is activated to enable manual intervention. The focal points in the scenario are the elevator's hook and the coke can's hanging ear, with all other information being superfluous to the background. Although the implementation of instance segmentation for detection purposes is viable, it is pertinent to note that existing instance segmentation models encounter numerous challenges:

1) The research is laser-focused on specific areas, but current models struggle to zoom in on these targets across the whole picture without getting bogged down by background details.

2) Dealing with tricky lighting is a big hurdle in image detection. The target's outdoor setting and around-the-clock operation mean coping with really varied lighting. Too much or too little light can wipe out key features of objects, and it also makes the background look more like the target, leading to mix-ups.

3) Limited datasets pose a challenge as they often fail to encompass the full spectrum of scenarios, particularly those that are extreme. This can lead to models making inaccurate predictions when faced with unfamiliar situations.

4) The dry quenching coke elevator operates with periodic changes in the hook and hanging ear status. During actions like hooking the ear, these changes occur rapidly. To keep up with these dynamic shifts and respond promptly, the detection system must operate at a high frequency. A slow-reacting system can't keep pace with real-time changes, risking misinterpretation of normal operations as faults. This not only triggers unnecessary shutdowns and hikes operational costs but might also obscure genuine issues, compromising both productivity and equipment safety.

5) Detecting detailed changes within the elevator's steel structure is crucial for monitoring its operational status. During the key phases of hook opening and closing, the internal gaps in the steel structure undergo minute alterations. These subtle changes are vital for assessing the condition of the steel structure. However, current instance segmentation models often fail to accurately capture these fleeting and essential features, leading to missed or misidentified characteristics. The misjudgment of these critical features can disrupt the accurate assessment of the elevator's operational status, potentially causing unexpected production halts.

These challenges make detection much harder and often result in missed details during actual inspections.

This paper presents an innovative instance segmentation framework, YOLOv11n-CL-seg, designed to address the challenges of uncertain illumination and dataset constraints through the implementation of an image data augmentation strategy. To enhance the detection of intricate details, the Local Region Self-Attention (LRSA) mechanism is incorporated to focus on localized target regions. To accelerate the detection process, the main network employs GhostConv in place of conventional Convolutional layers, and the Cross-scale Feature Fusion Module (CCFM) is utilized for efficient feature integration. These modifications result in a reduction of computational complexity and parameter count, thereby enhancing the detection speed.

III. Three Algorithm Improvements

A. Overall Network Structure

The revised YOLOv11n-seg features a streamlined network architecture, as depicted in Figure 2. It comprises an enhanced backbone network, a neck network, and the original model's detection head. Within the backbone, GhostConv replaces traditional Conv in layers 1, 3, 5, and 7. GhostConv innovates by creating "ghost feature maps" through unique linear transformations, which allows it to generate a diverse array of feature maps using fewer convolutional kernels. This approach not only reduces the parameter count but also boosts the network's efficiency[19]. It is proposed to substitute the C3k2 modules within the original model architecture with C3Ghost modules at the 4th, 6th, and 8th layers, with the objective of further diminishing the model's computational complexity[20]. Furthermore, the neck network incorporates the CCFM feature fusion method, a strategy that not only contributes to the model's lightweight design but also enhances its ability to discern and integrate meaningful information across multiple scales. This refinement significantly amplifies the model's proficiency in extracting features within complex scenes[21]. In response to the accuracy degradation associated with the model's optimization for lightweight performance, the system has implemented a strategic modification. Specifically, the C3k2 module, originally positioned at layers 15, 20, and 23 of the network, has been substituted with the C3Ghost-L module to enhance overall precision. The C3Ghost-L module is designed with an integration of the LRSA (Local Region Self-Attention) mechanism, enhancing the network's capability to discern pivotal local features and

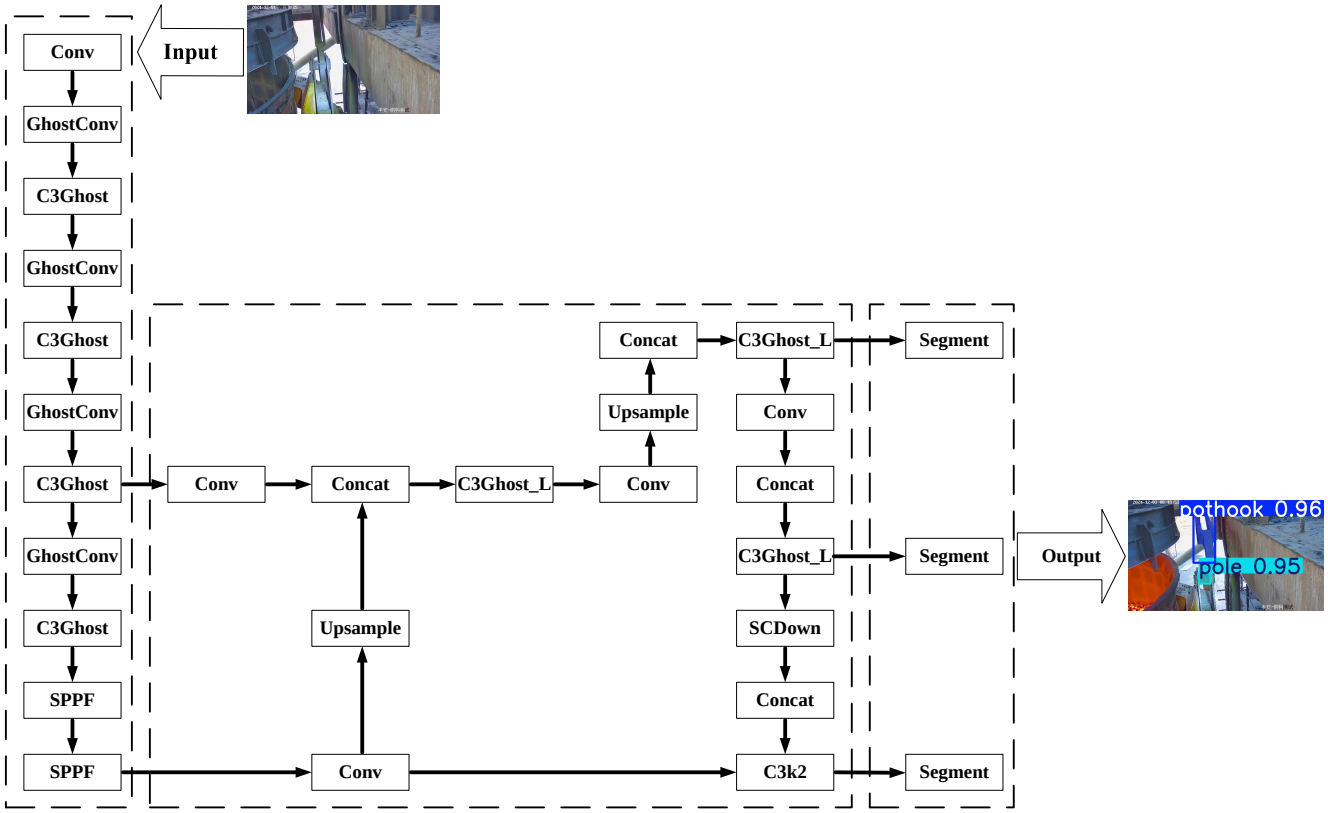

Fig. 2: YOLOv11n-CL-seg network structure

direct the system's focus towards essential areas. This design ensures the system sustains superior performance even under the constraints of lightweight optimization [22].

B. Main Network Improvements

In scenarios characterized by the constraints of embedded devices, the pivotal challenge lies in the optimization of fast inference, reduced energy consumption, and minimal memory usage for intelligent applications. This research introduces the Ghost module as a fundamental lightweight component, thereby enhancing the model's compatibility with embedded devices. The Ghost module employs a novel two-step convolutional approach to achieve lightweight architecture. In the initial phase, traditional convolutional operations are utilized to generate fundamental feature maps, thereby accomplishing preliminary feature extraction. Subsequently, linear transformations are applied to these base feature maps to produce an extended set of feature maps. This distinctive design effectively mitigates computational expense and complexity, enabling the model to execute feature extraction tasks with reduced resource consumption in embedded device environments. Figure3 provides a clear illustration of the lightweight operational mechanism and computational optimization rationale.

The Ghost module has given rise to the Ghost bottleneck layer, which is specifically engineered for smaller convolutional neural networks. This layer is constructed by cascading two Ghost modules. As depicted in Figure 4, when the stride is 1, the initial Ghost module expands the channel dimension to augment feature representation, while the subsequent module reduces the channel count to correspond with the output of the direct connection. The

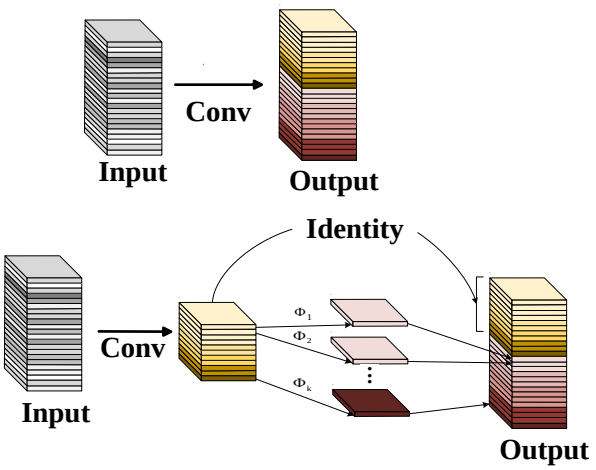

Fig. 3: Ghost Module

output of the first module undergoes batch normalization and ReLU activation to facilitate faster convergence and to intensify non-linearity. The second module employs batch normalization solely for the purpose of data stabilization. In cases where the stride is 2, the direct connection incorporates a downsampling layer to modify the data dimensions. Furthermore, a depthwise convolution with a stride of 2 is introduced between modules to bolster the extraction of multi-scale features. This structure also employs pointwise convolution to mitigate computational expenses. In the context of network optimization, this study implements a substitution of the conventional convolutional modules in layers 1, 3, 5, and 7 with the advanced GhostConv modules. Furthermore, it refines the bottleneck layer of the C3 module, transitioning it into a Ghost

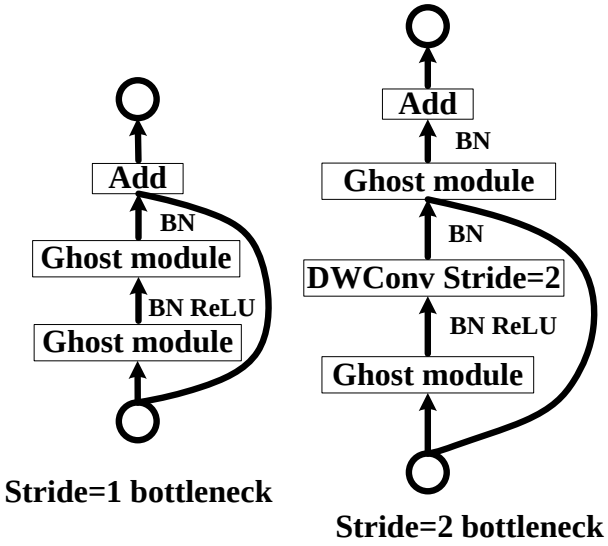

Fig. 4: Ghost bottleneck layer structure

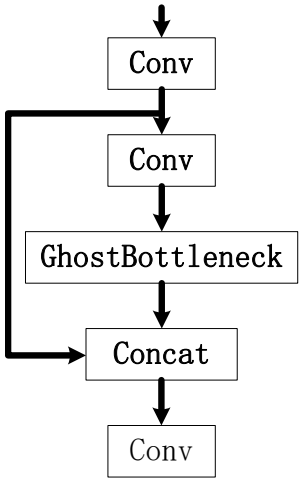

Fig. 5: C3Ghost

bottleneck layer, thereby constructing the C3Ghost module. This novel C3Ghost module is then utilized to replace the original C3K2 module. Referencing Figure 5, the replacement of the initial C3K2 module with the newly designed structure is illustrated. This design strategy is implemented to minimize computational load and memory usage, thereby optimizing the model's efficiency in operationally constrained settings, such as embedded environments.

C. CCFM Feature Fusion

In the neck network, the CCFM (Cross-Scale Feature Merger) method is implemented for feature fusion, comprising a series of structured operational units. The initial phase involves the adjustment of input feature channels and dimensions via a 1×1 convolutional layer. This is followed by upsampling to enhance the feature map's resolution. The subsequent step entails the integration of information across various scales through element-wise addition, facilitating the capture of contextual cues. After fusion, a 1×1 convolutional layer is employed for further refinement of the features. The process is finalized with additional upsampling and fusion, thereby reinforcing the multi-scale feature integration. By integrating upsampling and downsampling, the CCFM enables bidirectional fusion of features across different scales, effectively leveraging the informative content of each scale. The CCFM structure is illustrated in Figure 6.

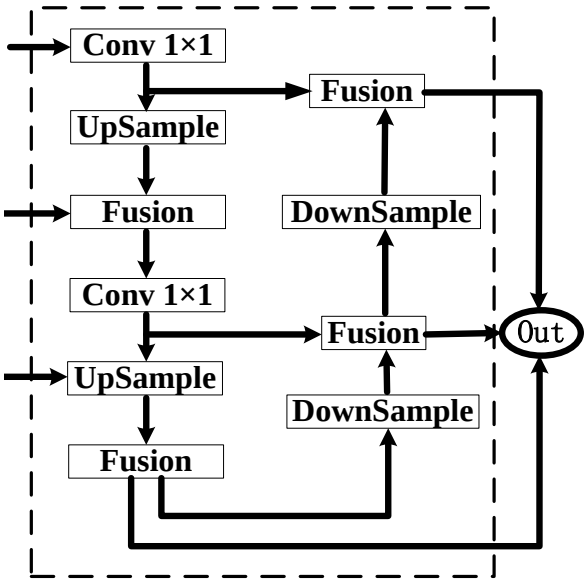

Fig. 6: CCFM Structural Diagram

D. LRSA Attention Mechanism

The current system is exclusively focused on enhancing elevator hooks and their associated components, making local information processing critical for its objectives. To augment the network's capacity to manage local feature intricacies, we have incorporated the LRSA (Local Region Self-Attention) mechanism. This mechanism comprises LayerNorm, WindowAttention, and ConvFFN, with WindowAttention serving as the central component. It segmentizes the input features into discrete local windows, executing self-attention computations within these confines. This strategy ensures a precise capture of local feature interdependencies while fostering information exchange across adjacent regions. The LRSA is specifically attuned to the feature extraction demands of hooks and ears. The ConvFFN is structured from convolutional and fully connected layers. Figure 7 delineates the LRSA module's architecture, elucidating the interconnections and data flow among its elements, and elucidating its operational methodology and performance advantages.

E. C3Ghost-L module

In order to enhance the model's performance and mitigate the accuracy degradation associated with its lightweight design, this research integrates the LRSA attention mechanism into the C3Ghost module, thereby constructing the C3Ghost-L module. The structure of this module is depicted in Figure 8. The C3Ghost-L module exhibits two principal advantages. Firstly, LRSA is adept at precisely identifying local feature dependencies, thereby enhancing the model's sensitivity to intricate details such as steel structures and ears. Secondly, LRSA employs a local window-based approach for attention calculation, which significantly reduces computational complexity. This method aligns exceptionally well with the lightweight design principles of C3Ghost. Within the C3Ghost-L module, the GhostBottleneck employs a two-step

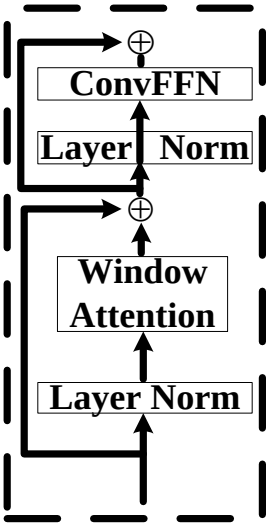

Fig. 7: LRSA structural diagram

convolution strategy to minimize computational load while generating comprehensive feature maps. Concurrently, LRSA focuses on facilitating local feature interaction, thereby delving deeper into the analysis of output features. The synergistic operation of these components further reduces computational demands. Moreover, the features processed by LRSA introduce an additional dimension to feature fusion, augmenting the model's cognitive and representational capacities, and effectively counteracting the accuracy loss incurred by the lightweight design.

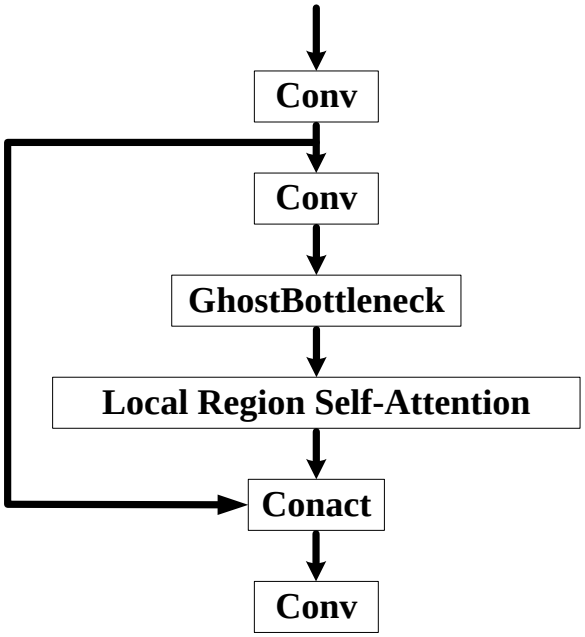

Fig. 8: C3Ghost-L Structural Diagram

IV. Experiment and Results Analysis

A. Experimental Dataset

This study built a dataset using 24-hour monitoring videos from Feng'an Steel Company's dry quenching coke production line. The video resolution is 2560*1440. Two frames per second were extracted from the collected videos, as shown in Figure 9. The suitable images were then manually selected and annotated.

The equipment comprises an HIKVISION outdoor surveillance camera, a fixed LED lighting system for nocturnal illumination, and an integrated digital video recorder (DVR) for the storage of acquired data. The support pole is strategically positioned at a fixed location, no more than 2 meters from the exterior of the elevator. The camera, mounted atop this pole, is angled to capture real-time imagery from above. As depicted in Figure 9, these data collectively form the dataset utilized in this study.

In the data preprocessing phase, this study segmented the collected data into seven distinct operational intervals, reflecting the continuous production cycle of a steel manufacturing facility. These intervals are categorized as follows: nighttime (00:00-05:00), dawn (05:00-08:00), morning shift (08:00-11:00), midday (11:00-13:00), afternoon (13:00-17:00), dusk (17:00-19:00), and evening (19:00-24:00). This classification system is designed to account for variations in lighting intensity, ambient temperature fluctuations, and equipment operational load, which are critical parameters in the production process. Utilizing an equal-interval sampling strategy, the study systematically captured the key operational stages of the elevator's working cycle, including descent, preparation for hook opening, hook opening, completion of hook opening, preparation for hook engagement, hook engagement, completion of hook engagement, and ascent. For each operational interval, 50 image samples were selected to provide a comprehensive representation of the process characteristics. The compiled dataset of 400 valid images was divided into training, validation, and test sets in a ratio of 7:2:1, ensuring an equitable distribution for model training and evaluation purposes.

In order to further improve the model's generalization capability, a variety of data augmentation techniques were implemented during the training phase. These techniques encompass random translation, random rotation, random grayscale conversion, and random cropping, among others. These augmentation procedures effectively emulate variations encountered in practical scenarios, thereby enhancing the model's robustness to fluctuations in lighting conditions, orientation, and scale.

B. Experimental Platform

The hardware and software environment for this experiment are as follows: The hardware setup includes a desktop computer running the Windows 11 operating system, equipped with an NVIDIA GeForce RTX 4060 Ti GPU (16GB GDDR6 VRAM), 32GB DDR5 RAM, and a 1TB NVMe solid-state drive. The software environment is built using Python 3.9.0, PyTorch 2.6.0, and the CUDA 12.6 framework.

To enhance the model's generalization capability, a multi-scale data augmentation strategy is adopted, which includes random translation ($\pm 15\%$ range), rotation (-30° to +30°), grayscale transformation (probability 0.2), and cropping (minimum retention rate of 0.7).

The training parameters are set as follows: 200 training epochs, a batch size of 32, an initial learning rate of 0.001, and the Adam optimizer (β_1=0.9, β_2=0.999) combined with cosine annealing learning rate scheduling (T_0=20). This scheduler helps the model avoid local optima and improves

Fig. 9: Dataset

generalization. The loss function used is the Dice-CE combined loss with a weight ratio of 1:1, balancing attention to foreground and background and improving instance segmentation accuracy.

These configurations contribute to enhancing the model's training speed, accuracy, and generalization capabilities, providing robust support for accurate status monitoring of the coke dry quenching lifting equipment.

C. Evaluation Metrics

The research selects mean Average Precision (mAP) to assess the overall performance of the model. Its calculation formula is as follows:

$$mAP = \frac{1}{N}\sum_{i=1}^{N}\int_{0}^{1} P(R)\,dR \times 100\% \tag{1}$$

$$P = \frac{TP}{TP + FP} \tag{2}$$

$$R = \frac{TP}{TP + FN} \tag{3}$$

In this context, P denotes precision and R denotes recall. Recall is defined as the ratio of actual positive samples within the total predicted samples. TP (True Positives) refers to the count of instances that are correctly classified as positive, specifically those instances that are genuinely positive and labeled as such by the classifier. FP (False Positives) represents the count of instances incorrectly classified as positive, encompassing instances that are actually negative but misclassified as positive by the classifier. FN (False Negatives) indicates the count of instances incorrectly classified as negative, referring to instances that are truly positive but erroneously labeled as negative by the classifier.

The degree of model lightweighting is measured by three indicators: the number of parameters, the amount of floating-point operations, and the size of model weights; FPS represents the number of images (frames) that can be processed per second, which is used to assess the real-time segmentation speed of the model.

D. Comparative Experiment

In order to conduct a comprehensive and unbiased evaluation of the YOLOv11n-CL-seg model's performance advantages in the context of coke quenching elevator image segmentation, a comparative analysis was performed with a selection of leading instance segmentation models, namely Mask R-CNN, YOLACT, BlendMask, YOLOv5-seg, YOLOv8n-seg, YOLOv9c-seg, and YOLOv12n-seg. The experiment was conducted under uniform conditions, utilizing the same dataset and hyperparameter configurations (including a learning rate of 0.001, batch size of 16, and 80 epochs) on an RTX4060Ti16GB graphics card. To ensure the integrity and reliability of the comparative results, the training process was replicated five times. The experimental setup included CUDA 12.6, Python 3.9, and PyTorch 2.6.0, with all training strategies consistent with the current study.

Table 1 shows how various models did on the task of segmenting images of coke dry quenching hoists. The YOLOv11n-CL-seg model we came up with in this study really did a great job, outperforming popular instance segmentation models like Mask R-CNN, YOLACT, BlendMask, YOLOv5-seg, YOLOv8n-seg, YOLOv9c-seg, and YOLOv12n-seg across the board. To be more specific, the YOLOv11n-CL-seg model really shines when it comes to important things like how accurate it is at drawing the boxes around objects (precision), how good it is at finding

TABLE I: COMPARISON OF DIFFERENT MODELS

Module	box			mask			Params/M	FPS(frame $\cdot s^{-1}$)
	Precision	Recall	mAP@0.5	Precision	Recall	mAP@0.5		
Mask R-CNN	0.881	0.863	0.952	0.877	0.831	0.947	46.4	42.7
YOLACT	0.874	0.852	0.931	0.865	0.814	0.924	32.1	53.4
Blend Mask	0.891	0.872	0.957	0.884	0.837	0.954	54.6	38.5
YOLOv5-seg	0.912	0.884	0.961	0.891	0.852	0.958	2.51	96.1
YOLOv8n-seg	0.986	0.973	0.981	0.983	0.971	0.979	2.94	91.1
YOLOv9c-seg	0.971	0.964	0.968	0.965	0.961	0.963	23.67	67.4
YOLOv11n-seg	0.993	0.982	0.985	0.989	0.98	0.982	2.84	94.2
YOLOv12n-seg	0.979	0.968	0.978	0.973	0.965	0.971	2.82	93.1
YOLOv11n-CL-seg	0.991	0.98	0.983	0.99	0.976	0.981	1.47	120.5

TABLE II: RESULTS OF ABLATION STUDY PERFORMED

Order	GhostConv	C3Ghost	CCFM	C3Ghost-L	mAP@0.5		Params/M	FLOPs/G	Weight(MB)
					box	mask			
1					0.985	0.982	2.84	10.4	5.73
2	✓				0.971	0.970	2.60	9.7	5.29
3	✓	✓			0.961	0.959	2.26	9.0	4.66
4	✓	✓	✓		0.956	0.953	1.41	8.0	3.16
5	✓	✓	✓	✓	0.983	0.981	1.47	8.2	3.21

all the objects it should (recall), and its overall average performance score (mAP) for both the boxes and the detailed outlines (masks). The data demonstrates that the model exhibits high precision in object boundary localization, achieves comprehensive object identification, and delivers segmentation results for object contours that align closely with actual conditions. Although the YOLOv11n-CL-seg model introduced in this research exhibits a marginal decrease in accuracy relative to the YOLOv11n-seg model, it demonstrates a substantially lower parameter count compared to both YOLOv11n-seg and the other benchmarked models. A comprehensive assessment indicates that the YOLOv11n-CL-seg model developed in this research fulfills the accuracy requirements for industrial scene detection. Concerning real-time performance, the model achieves an FPS value of 120.5 frames per second. The model's actual detection response time demonstrates significantly superior performance compared to alternative models. Consequently, the proposed model offers highly efficient technical support for the dynamic monitoring of industrial equipment.

E. Ablation Experiment

To precisely dissect the specific contribution of each module to the overall performance in the improved YOLOv11n-CL-seg model, module-level ablation studies were conducted. A variety of variant models were constructed by sequentially removing or replacing certain modules based on the fully improved model, and tested under the same experimental dataset and environment. The results are shown in Table 2. The four improved modules included GhostConv, C3Ghost, CCFMand C3Ghost-L.

Based on the results of the ablation experiment, the baseline model (sequence number 1) showed initial performance, with mAP@0.5(box) of 0.985, mAP@0.5(mask) of 0.982, 2.84M parameter volume, and FLOPs of 10.4G. When gradually introducing improved modules, the performance exhibits different changes. After introducing the GhostConv module (sequence number 2), the model parameter volume and FLOPs decreased somewhat, but the accuracy also decreased, with mAP@0.5(box) dropping to 0.971 and mAP@0.5(mask) dropping to 0.970, indicating that this module affects accuracy while being lightweight. Further introduction of the C3Ghost module (sequence number 3), both the parameter volume and computation volume continue to decrease, but the loss of accuracy is exacerbated. The addition of the CCFM module (sequence number 4), significantly reduces the parameter volume to 1.41M and FLOPs to 8.0G, but the drop in accuracy is significant. When all improved modules including the C3Ghost-L module (sequence number 5) are introduced, the model achieves a good balance of performance and resource consumption while maintaining its lightweight advantage, with mAP@0.5(box) reaching 0.983 and mAP@0.5(mask) reaching 0.981, which are close to the accuracy of the baseline model. This highlights the critical importance of the synergy of various modules for model optimization. Please help me translate.

V. Fault Model Construction for Coke Bucket Hoist

A. State Analysis of Coke Bucket Hoist

Instance segmentation technology is employed to precisely isolate the components of a hoist amidst complex backgrounds through the identification of spatial locations and frame motion instance features. Nonetheless, a solitary

static image is limited to capturing a fleeting state, insufficient for depicting the full operational dynamics of the hoist or serving as a dependable criterion for process validation. The hoist is susceptible to three predominant faults: anomalous hook descent during the discharge of empty coke drums, premature hook engagement when the ear has not attained the designated zone, and deviations in hook engagement position. The detection of these faults necessitates the analysis of temporal attributes, dynamic procedures, and sequential actions, rendering a single frame image inadequate for the task.

The present study introduces an innovative method to address the challenges outlined. This method leverages the feature-rich information derived from instance segmentation techniques. In the operation of the hoist, periodic gaps within the steel framework are discernible, represented as black areas enclosed within the blue structure in the segmentation mask image (refer to Figure 10 for visualization, where the hook is depicted in blue and the lug in cyan). Furthermore, the instance segmentation mask image exhibits a unique pattern of change in the number of lugs. During the no-load phase, typically, a single lug is visible. However, during the coke drum grasping action, the mechanical occlusions cause the single lug to be segmented as two in the mask. Utilizing this information, the hoist's operational process is delineated into eight distinct time periods. By integrating the image data and characteristic features of each period, a comprehensive, multi-dimensional dynamic analysis system is constructed. This system empowers computers to make intelligent and precise judgments regarding the hoist's operational status, significantly enhancing the timeliness and accuracy of fault detection.

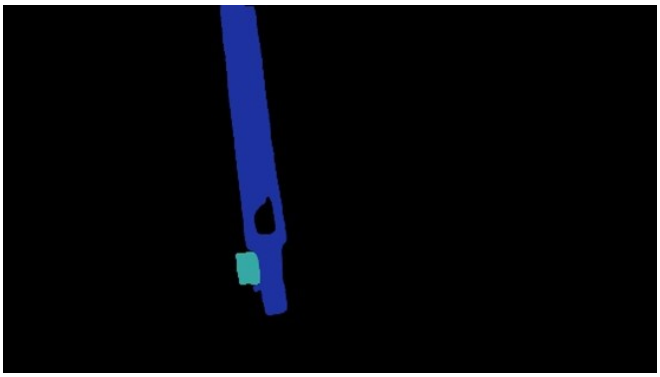

Fig. 10: Instance Segmentation Mask Map

B. Division and Qualitative Analysis of Lifting Machine Working States

The precise differentiation of operational phases in the dry quenching coke can hoist is essential for the effective surveillance of its operational status. This research employs instance segmentation technology, mask feature analysis, and encoder information from the hoist to accurately demarcate the various operational states. The classification of these states is based on an extensive statistical analysis of operational data gathered from real-world applications. The study has involved prolonged monitoring of the hoist, compiling data on the operational durations of key actions such as opening and closing hooks, as well as during different operational stages. The findings indicate a high level of stability and regularity in these actions and stages under normal operational conditions. Hence, the temporal parameters for each phase in this study are derived from these statistical data, ensuring a scientifically rigorous and accurate categorization of time periods. This data provides a solid foundation for the precise monitoring of the hoist's operational status. Details are as follows:

Descending phase: The encoder installed on the hoist is designed to provide positional feedback. In instances where the encoder registers a consistent downward movement of the hoist, this period is classified as the hoist's descending phase.

Hook opening preparation phase: Upon reaching the lowest point as indicated by the encoder feedback, a transition in the number of hanging ears from two to one is observed through instance segmentation. Concurrently, the mask map depicts the emergence of a black area within the steel structure, signifying the hoist's transition to the preparation phase for hook opening. As shown in Figure 11, this phase is consistently observed to have a duration of 4 seconds. This duration has been established through a comprehensive statistical analysis of extensive operational data of the hoist. Long-term monitoring has indicated that under standard operating conditions, the duration of this phase remains stable at approximately 4 seconds, thus serving as the reference for time setting in this phase.

Duration of the hook opening process: The analysis of the instance segmentation mask map reveals a sequential increase, followed by a minor decrease, and eventual stabilization of the internal gap area within the steel structure. This period is classified as the hoist's hook opening process. As shown in Figure 12, this phase lasts for 9 seconds.

Hook Completion Phase: The analysis of the mask diagram indicates a consistent internal gap area within the steel structure, sustained for a period of one second. This duration is classified in the research as the phase of completion for the hoist's hook opening process.As depicted in Figure 13, this phase of completion is observed to last for one second.

Upon the hanging ear's arrival at the pre-established ROI area, the hook initiates its descent. Simultaneously, the mask diagram indicates a progressive enlargement of the internal gap area within the steel structure. This period is designated within the study as the preparation phase for the hoist's grappling mechanism. As shown in Figure 14, this phase lasts for 4 seconds.

Hook Engagement Phase: During the continuous descent of the hook, the mask diagram indicates a progressive decrease in the internal gap area of the steel structure. This period is classified within the study as the phase during which the hoist is engaged in the act of grasping. As shown in Figure 15, this phase lasts for 6 seconds.

Gripper Completion Period: The mask map indicates the disappearance of internal voids within the steel framework. Instance segmentation has identified a transition from a single hanging ear to two. This period is classified as the completion phase of the hoist's grappling operation. As shown in Figure 16, this period lasts for 5 seconds.

Ascension Period: The encoder has observed a sustained

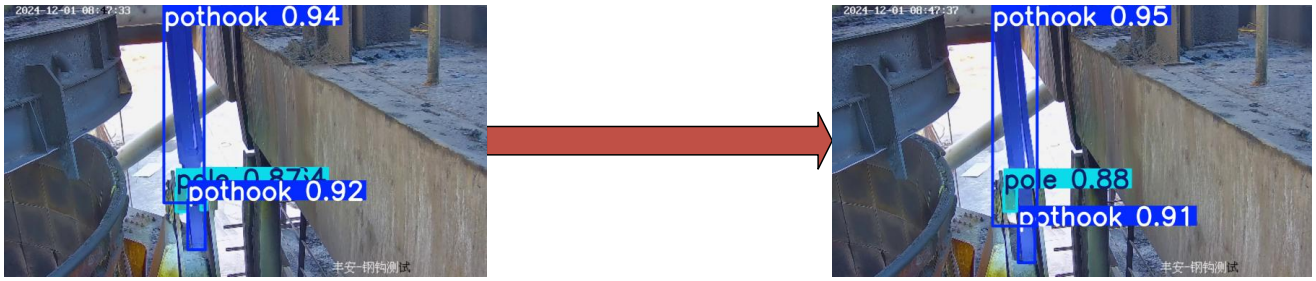

Fig. 11: Instance Segmentation Mask Map

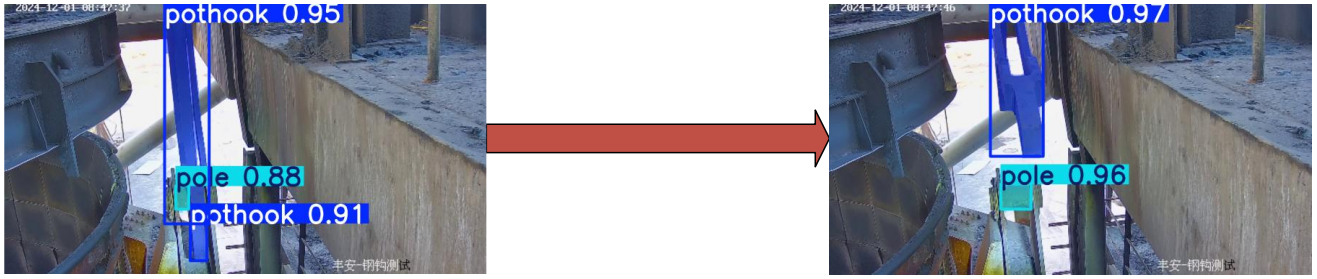

Fig. 12: Instance Segmentation Mask Map

Fig. 13: Instance Segmentation Mask Map

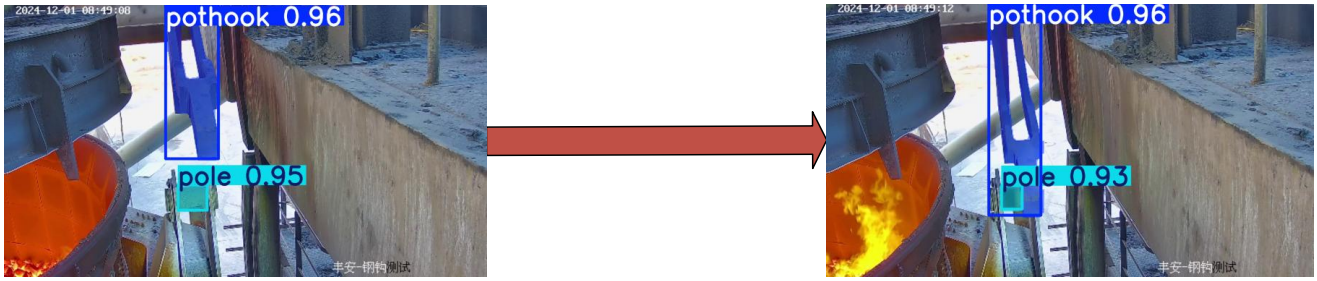

Fig. 14: Instance Segmentation Mask Map

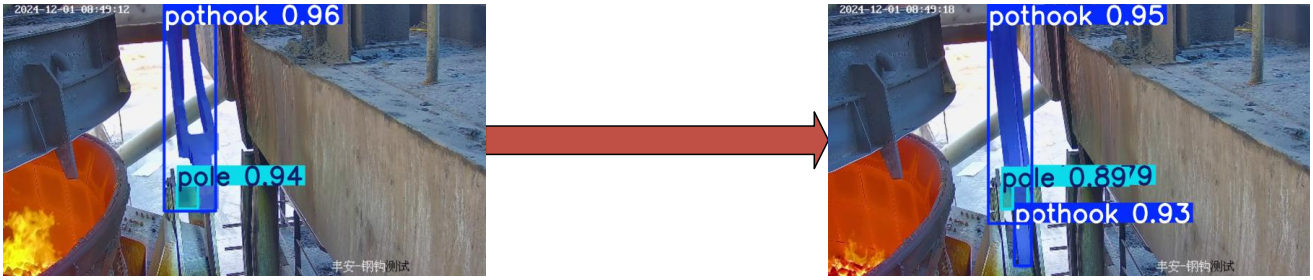

Fig. 15: Instance Segmentation Mask Map

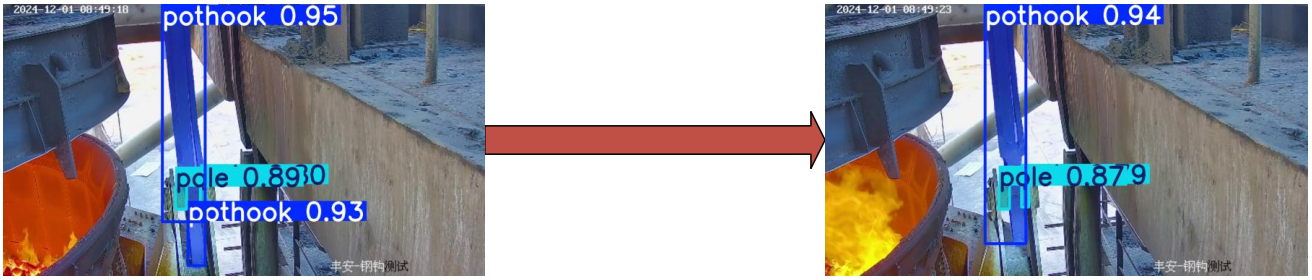

Fig. 16: Instance Segmentation Mask Map

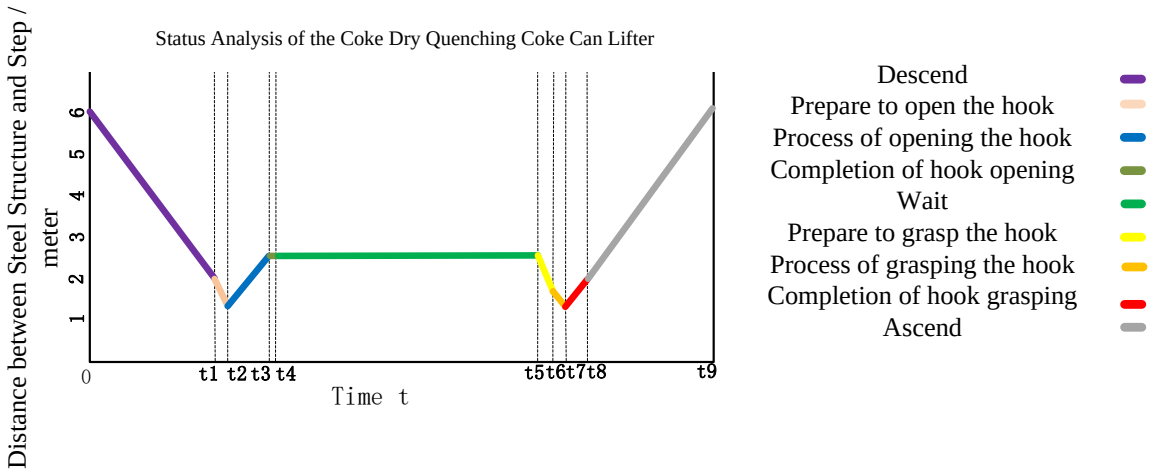

Fig. 17: Instance Segmentation Mask Map

upward trend in the hoist's location. This period is classified as the hoist's ascending phase. Upon reaching this stage, the hoist has accomplished a full workflow cycle.

C. Design of Hoist Monitoring System

The current study elucidates a novel methodology for the surveillance of dry quenching coke can hoists. Central to this methodology is the utilization of temporal constraints across various intervals, coupled with the assurance of operational process integrity for anomaly detection. In normal operational circumstances, the hoist's sequential stages must adhere rigorously to predefined temporal parameters, with each stage exhibiting a full spectrum of critical actions and state transitions. As shown in Figure 17, the operational phases are delineated as follows: 0-t1 for descent, t1-t2 for hook opening preparation, t2-t3 for the hook opening process, t3-t4 for hook opening completion, t4-t5 for a waiting interval, t5-t6 for hook closing preparation, t6-t7 for the hook closing process, t7-t8 for hook closing completion, and t8-t9 for ascent. Should the surveillance system ascertain that the actual operational duration of a specific phase transcends the established temporal threshold and fails to progress seamlessly into the subsequent phase, it is indicative of operational stagnation or malfunction during said phase. Promptly, the system activates an alarm mechanism, facilitating expeditious intervention to rectify the issue, thereby safeguarding the continuity and stability of the dry quenching coke production process. Furthermore, the continuous monitoring of operational integrity across each phase significantly amplifies the system's preemptive capabilities in identifying potential hoist failures. This encompasses the surveillance of key components' action states, lug quantity and positional variations, and the dynamic evolution of mask map features. This multifaceted monitoring approach ensures the secure and efficient operation of the dry quenching coke process.

D. Industrial Scenario Verification

The dry quenching coke production line at Ansteel Iron and Steel Company has successfully implemented a deep learning-based monitoring system, which underwent a continuous one-month operational test. Throughout the testing phase, the system demonstrated robust performance with no interruptions due to software malfunctions. It effectively exchanged data with other equipment control systems to acquire real-time operational data of the coke can hoist, enabling detailed analysis and comparison. The system's ability to precisely track the hoist's operational status remained consistent, irrespective of the production load, thereby ensuring stable detection performance under varying operational conditions.

The system incorporates three key mechanisms, which are continuous state evaluation, time-duration limitations, and historical data analysis, to ensure comprehensive monitoring of the hoist. In the context of monitoring state transitions of the hoist, the enhanced model necessitates three consecutive and consistent detection results for confirmation. Leveraging a foundational accuracy rate of 98.1% and prioritizing the detection of intricate details within steel structures, the model effectively resolves the steel structure detail detection issue posed in Chapter 2. As shown in Figure 18, the enhanced model successfully mitigates the risk of erroneous state interpretations resulting from the abnormal disappearance of black areas within the steel structure on the mask graph. With a frame rate of 120.5 frames per second (fps), this model exhibits superior responsiveness and stability under complex operational conditions, outperforming alternative models. The system's enhanced precision in monitoring the steel structure phase is attributed to the improved model's higher fps. It automatically retrieves and compares data from the preceding eight work cycles during the monitoring process. In the event of an anomaly, the system undergoes a secondary review. The system's ability to more accurately correct deviations is a direct consequence of the improved model's higher initial accuracy compared to other models. These three pivotal mechanisms substantially enhance the reliability of the monitoring system within complex industrial environments.

VI. System Overview

A. System Architecture

The system incorporates a camera for the acquisition of images and employs the OpenCV library for the decoding of video streams. It is designed to process two frames per

(a) Improved Model Detection

(b) Other Model Detection

Fig. 18: Detailed inspection diagram of the hook grasping process

second for semantic segmentation tasks. The comprehensive system flow is depicted in Figure 19.

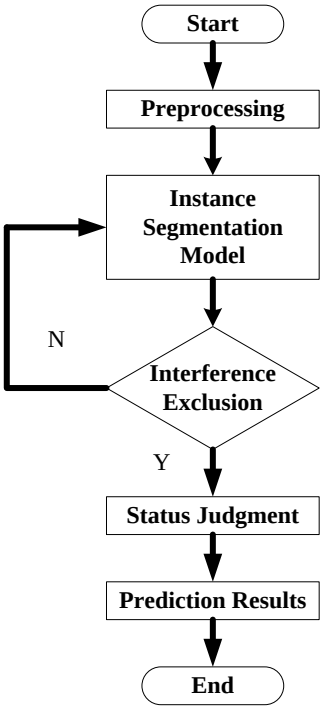

Fig. 19: System Workflow Diagram

1) The initial phase involves the application of traditional machine vision techniques for the preprocessing of acquired images. Subsequently, these processed images are input into an instance segmentation model for the detection process.

2) The system integrates the analysis of state changes to ascertain the condition of the coke can hoist. It evaluates the presence of significant disturbances in the operational environment based on the detection status. In the event of detected interference, the system activates a delayed response. This process leverages the mechanisms of continuous state evaluation, state time constraints, and historical data retrospective analysis to guarantee the precision and reliability of the judgments.

3) The system, developed utilizing LabVIEW software, incorporates a comprehensive human-machine interface. This interface facilitates real-time monitoring of the coking drum elevator's operational status and segmentation outcomes. In the event of an anomaly, the system is designed to activate an immediate alarm. This integrated solution is pivotal in ensuring the prompt detection and resolution of potential faults in the elevator, thereby enhancing overall system reliability and safety.

B. Human-Machine Interface

The system is designed to provide timely information regarding the status of the dry quenching coke drum elevator. The login interface requires users to enter their credentials and passwords to ensure secure access. Once logged in, users are presented with a detection interface that includes a detailed curve chart. This chart visually represents the elevator's operational status, with the horizontal axis indicating various states and the vertical axis denoting positions. In instances where the elevator's operational cycle is within acceptable parameters, the detection system remains in its standard operational state. Conversely, if the cycle deviates from the established norms, the system enters a cautionary state, initiating an abnormal alarm procedure. The surveillance video feed offers a direct visual assessment of the coke drum elevator's condition, accompanied by a set of indicator lights and alarm signals located beneath the display. The parameter configuration interface is primarily utilized for adjusting the alarm threshold settings. The alarm log function is designed to document all instances of warnings and irregularities pertaining to the elevator, with these records being securely stored for subsequent inquiry. Figure 20 provides a schematic illustration of this setup.

The human-machine interface system integrates the following functional modules:

1) Login interface: User verification is conducted through password-protected access.

2) Detection interface: Work detection curve for the coke bucket hoist. Normal: The coke bucket hoist's work cycle is normal. Warning: The hook warning is triggered when the lug position does not reach the designated ROI area and the hoist enters the grab hook state. Abnormal: An exception is triggered if the coke bucket hoist's work cycle is abnormal. All statuses are visually represented using indicator lights. If a light is on, it signifies that the system is in the corresponding state.

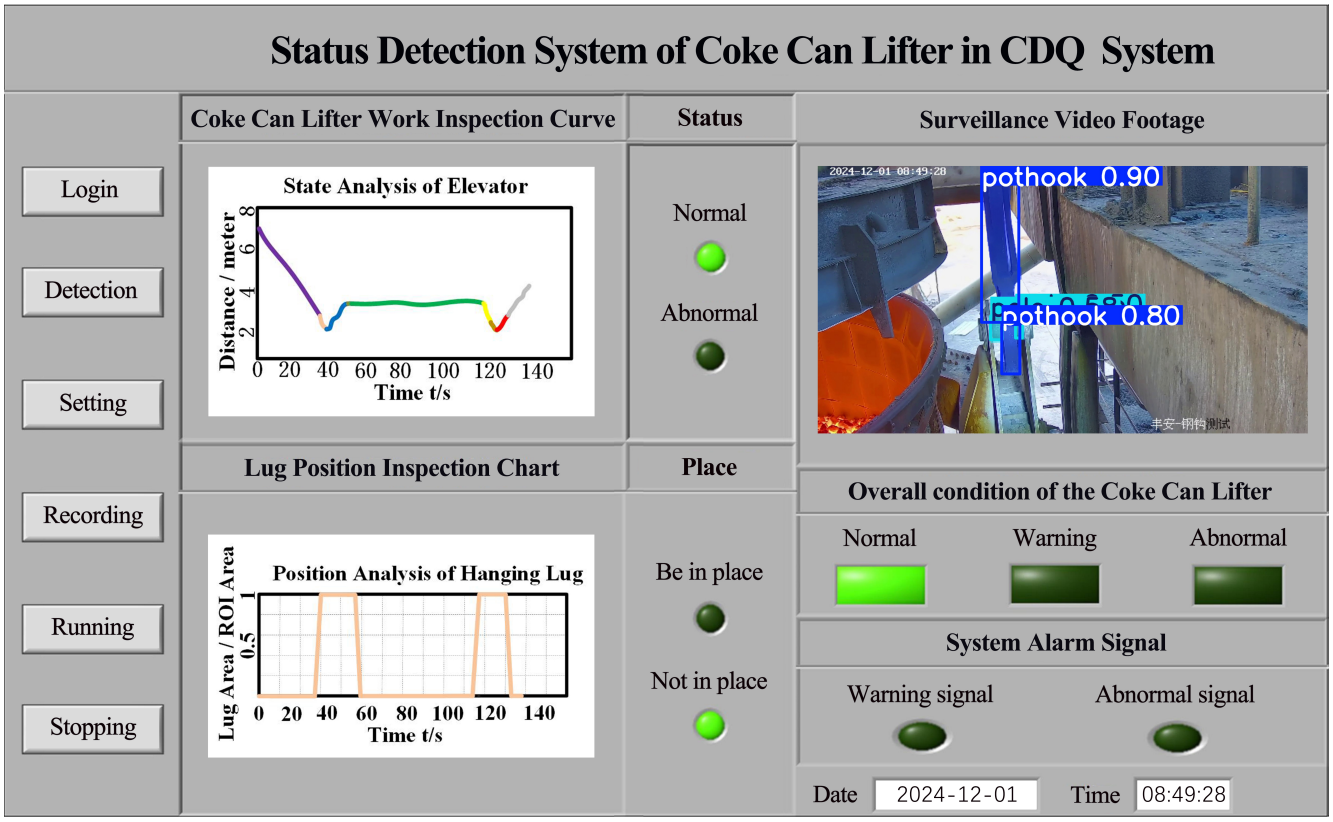

Fig. 20: System Workflow Diagram

3) Real-time Monitoring: Real-time display of the surveillance area's visuals.

4) Configuration Panel: Adjustable alarm thresholds and lag parameters.

5) Alarm records: Timestamp records of operational anomalies.

6) Start and Stop System: Controls the normal operation and stoppage of the detection system.

VII. Conclusion

The present study has successfully developed an advanced deep learning-based monitoring system for dry quenching hoists, substantially mitigating the challenges associated with conventional monitoring methodologies. Through a comprehensive analysis of the hoist's operational dynamics and the categorization of its various states, the system achieves high precision in recognition utilizing state-of-the-art instance segmentation techniques. In response to the identified limitations of the YOLOv11n-seg model, an innovative enhancement, the YOLOv11n-CL-seg model, has been proposed. This model demonstrates significant improvements in terms of lightweight design and operational efficiency. It maintains detection accuracy comparable to the original model while reducing the number of parameters, memory consumption, and computational requirements to 51.7%, 56%, and 78.8% of the original model, respectively. With a frame rate of 120.5 FPS, the model is highly suitable for real-time industrial monitoring applications. The experimental dataset, augmented through advanced data augmentation techniques, highlights the model's superior performance in accuracy, training efficiency, and complexity management. The YOLOv11n-CL-seg model's metrics for box and mask in terms of Precision, Recall, and mAP@0.5 are nearly equivalent to those of the YOLOv11n-seg model, yet it boasts significantly lower Params and a considerably higher FPS. The system's employment of the YOLOv11n-CL-seg model as the instance segmentation model ensures high accuracy and rapid response capabilities. It guarantees accurate identification of the hoist's operating state through continuous state evaluation and historical data review, thus ensuring production safety and reliability. The system has been successfully integrated into the production line of Feng'an Steel Company. The LabVIEW-based human-machine interface facilitates real-time monitoring, segmentation result display, abnormal alarm notification, and information recording, substantially enhancing the system's user experience and practical utility. This study represents a significant advancement in the field of intelligent hoist monitoring, with the developed methods and models holding extensive potential for future applications. Future research directions may include the integration of the Internet of Things and big data analytics for comprehensive intelligent production management, as well as the continuous optimization of models to enhance their performance in complex industrial environments.

References

[1] Vijay Bhatt, Manos De, Bulu Chandra, Bratindra Narayan Dey, Sphatick Bhattacharyya, Sujit Kumar Halder and Sharat Kumar, "Design and Engineering Challenges for Installation of Coke Dry Quenching Plant in Brown Field Scenario," Transactions of the Indian Institute of Metals, vol. 74, no. 5, pp1141–1153, 2021.

[2] Tianfei Zhou and Wenguan Wang, "Cross-Image Pixel Contrasting for Semantic Segmentation," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 46, no. 8, pp5398–5412, 2024.

[3] Qing Yu, Xinyu Ouyang, Bochao Su, Nannan Zhao, and Hongman You, "Vehicle Detection Algorithm in Complex Scenes Based on Improved YOLOv8," IAENG International Journal of Computer Science, vol. 52, no. 4, pp886-893, 2025.
[4] Lamin L. Janneh, Youngjun Zhang, Mbemba Hydara and Zhongwei Cui, "Deep Learning-Based Hybrid Feature Selection for the Semantic Segmentation of Crops and Weeds," ICT Express, vol. 10, no. 1, pp118–124, 2024.
[5] Ross Girshick, Jeff Donahue, Trevor Darrell amd Jitendra Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," Proceedings of the IEEE conference on computer vision and pattern recognition 2014, 24-27 June, 2014, Columbus Ohio, USA, pp580-587.
[6] Joseph Redmon, Santosh Divvala, Ross Girshick and Ali Farhadi, "You only look once: Unified, real-time object detection," Proceedings of the IEEE conference on computer vision and pattern recognition 2016, 27-30 June, 2016, Las Vegas Nevada, USA, pp779-788.
[7] Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov and Sergey Zagoruyko, "End-to-end Object Detection with Transformers," Lecture Notes in Computer Science: Proceedings of The European Conference on Computer Vision 2020, 23-28 August, 2020, Glasgow, U.K., pp213-229.
[8] Ao Wang, Hui Chen, Lihao Liu, Kai Chen, Zijia Lin, Jungong Han and Guiguang Ding, "Yolov10: Real-time end-to-end object detection," arXiv preprint arXiv:2405.14458, 2024.
[9] Xiaoliang Zhu, and Ziwei Zhou, "Infrared Action Recognition Model Based on Improved ST-GCN," IAENG International Journal of Computer Science, vol. 52, no. 6, pp1664-1671, 2025.
[10] Haiying Qi, Quanyan Gao, and Shulong Wang, "A License Plate Detection and Recognition Model in Complex Scenarios," IAENG International Journal of Computer Science, vol. 52, no. 6, pp1795-1805, 2025.
[11] Ronneberger Olaf, Fischer Philipp and Brox Thomas, "U-net: Convolutional networks for biomedical image segmentation," Lecture Notes in Computer Science: Proceedings of Medical Image Computing and Computer-Assisted Intervention 2015, 5-9 October, 2015, Munich, Germany, pp234-241.
[12] Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang and Jiaya Jia, "Pyramid Scene Parsing Network," Lecture Notes in Computer Science: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2017, 21-26 July, 2017, Honolulu, USA, pp2881-2890.
[13] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C. Berg, Wan-Yen Lo, Piotr Dollár and Ross Girshick, "Segment Anything," Lecture Notes in Computer Science: Proceedings of the IEEE/CVF International Conference on Computer Vision 2023, 1-6 October, 2023, Paris, France, pp4015-4026.
[14] Yun Bai, En Lu, and Hao-bo Wang, "A Pavement Crack Segmentation Algorithm Based on I-U-Net Network," IAENG International Journal of Computer Science, vol. 52, no. 6, pp1833-1844, 2025.
[15] Kaiming He, Georgia Gkioxari, Piotr Dollár and Ross Girshick, "Mask R-CNN," Lecture Notes in Computer Science: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2017, 21-26 July, 2017, Honolulu, USA, pp2961-2969.
[16] Xinlong Wang, Tao Kong, Chunhua Shen, Yuning Jiang and Lei Li, "SOLO: Segmenting Objects by Locations," Lecture Notes in Computer Science: Proceedings of the 16th European Conference on Computer Vision (ECCV) 2020, 23-28 August, 2020, Glasgow, U.K., pp649-665.
[17] Bowen Cheng, Ishan Misra, Alexander G. Schwing, Alexander Kirillov and Rohit Girdhar, "Masked - attention Mask Transformer for Universal Image Segmentation," Lecture Notes in Computer Science: Proceedings of the Conference on Computer Vision and Pattern Recognition 2022, 18-24 June, 2022, New Orleans, USA, pp1290-1299.
[18] Xu Tan, and Ji Zhao, "Multi-Task Vehicle Surface Damage Analysis Model Based on YOLOv8," IAENG International Journal of Computer Science, vol. 52, no. 6, pp1921-1929, 2025.
[19] Kai Han, Yunhe Wang, Qi Tian, Jianyuan Guo, Chunjing Xu and Chang Xu, "GhostNet: More Features from Cheap Operations," Lecture Notes in Computer Science: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 2020, 13-19 June, 2020, Seattle, USA, pp1580-1589.
[20] Mohammad Farid Naufal and Selvia Ferdiana Kusuma, "BloodCell-YOLO: Efficient Detection of Blood Cell Types Using Modified YOLOv8 with GhostBottleneck and C3Ghost Modules," Journal of Information Systems Engineering and Business Intelligence, vol. 11, no. 1, pp41–52, 2025.
[21] Yian Zhao, Wenyu Lv, Shangliang Xu, Jinman Wei, Guanzhong Wang, Qingqing Dang, Yi Liu and Jie Chen, "Detrs beat yolos on real-time object detection." arXiv preprint arXiv:2304.08069, 2024.
[22] Xiang Shen, Dezhi Han, Zihan Guo, Chongqing Chen, Jie Hua and Gaofeng Luo, "Local Self-Attention in Transformer for Visual Question Answering," Applied Intelligence, vol. 53, no. 13, pp16706–16723, 2022.