Research on State Detection of Ore Crusher Inlet Based on U-Net Network

Chengcheng Xu, Shaochuan Xu, Minghao Ma

Abstract—**To address the challenges of monitoring ore crusher inlets in complex, open-air environments, this study integrates image processing techniques with deep learning, proposing an enhanced system based on an improved U-Net algorithm. The system is designed to mitigate issues like clogging, improving operational efficiency. Image preprocessing reduces interference from variable lighting and dust. Incorporating channel attention (SE) and multiscale dilated convolution attention (MDCA) modules enhances segmentation accuracy by capturing local details and global context while filtering irrelevant features. Haar wavelet downsampling preserves essential structural information during feature compression. Empirical validation using real-world data shows the system's effectiveness in condition monitoring via non-continuous video frames. Performance evaluation indicates a mean Intersection over Union (mIoU) score of 92.6%, demonstrating high precision and reliability. Field deployments confirm the system's practical value in preventing equipment failures and optimizing mineral processing efficiency.**

Index Terms—**Industrial Image Inspection,crusher, semantic segmentation,image enhancement.**

I. Introduction

IRON ore, as one of the fundamental raw materials in modern industry, undergoes processing procedures that typically include mining, crushing, and grinding. Among these operations, the crushing process plays a particularly crucial role in breaking large ore blocks into appropriately sized fragments. However, in practical operations, irregular ore shapes often lead to frequent clogging incidents , causing crusher malfunctions that disrupt production efficiency and potentially damage equipment while increasing maintenance costs. Therefore, effectively monitoring crusher status and promptly addressing blockages has become critical for improving iron ore processing efficiency.

Traditional manual observation methods for Crusher inlet not only impose substantial labor costs through multi-operator requirements, but also introduce miscalculations caused by operator fatigue during extended monitoring periods. The evolution of computer vision technology has spurred development of conventional image processing-based detection methodologies, with notable progress achieved through watershed algorithms, threshold segmentation techniques, and support vector machine (SVM) based solutions . However, these conventional approaches demonstrate limited adaptability in complex operational environments. Critical limitations emerge when handling substantial variations in ore characteristics between production batches particularly or when confronting positional deviations of surveillance cameras that alter the geometric perspective of crushing chamber openings. Such scenarios necessitate cumbersome parameter recalibrations and restrict the generalizability of these methods across diverse industrial applications. Modern computer vision techniques, especially semantic segmentation models, leverage deep learning architectures to show stronger robustness and adaptability in varying environments and equipment positions.

Semantic segmentation technology has become the most groundbreaking image parsing method in the field of computer vision by realising accurate machine parsing of image content through pixel-level semantic annotation. Currently, three mainstream technology systems are formed in this field: U-Net [1], PSP-Net [2] and DeepLabv3+ [3] architecture [4]. Studies have shown that U-Net demonstrates unique advantages in industrial scenarios by virtue of its multi-level feature fusion mechanism: firstly, its symmetric coding and decoding structure can effectively capture the fine-grained texture features of the ore image; secondly, the hopping connection design significantly improves the transfer efficiency of the edge features; and ultimately, the parameter sharing mechanism achieves a stable performance in small-sample scenarios. Taking the research of Wang's team as an example, their innovative work includes: constructing the first multi-task U-Net framework for ore image segmentation and developing a new U-Net architecture with boundary detection; to confirm the method's significant advantages in industrial quality inspection scenarios[5].

Nevertheless, practical segmentation tasks often encounter challenges in defining boundaries between targets and backgrounds, which basic U-Net struggles to resolve effectively. Consequently, many researchers have attempted to integrate U-Net with other networks for dual-network detection and segmentation. For example, Chicchon M et al. compared U-Net with DeepLab-V3 performance and proposed a method combining image contour features with compound loss functions to achieve high-precision underwater image segmentation [6]. Liu et al. developed an improved Faster-SCNN and channel attention mechanism-based U-Net network for ore images, balancing segmentation accuracy with computational efficiency [7].

However, such detection strategies typically require coordination between multiple algorithms with varying processing speeds, necessitating synchronization of inference times during implementation and resulting in operational complexity. Additionally, model considerations must account for detection speed, stability, and robustness. Chen et al.

Manuscript received March 10, 2025; revised May 31, 2025.

Chengcheng Xu is a Graduate Student of School of Electronic Information, Liaoning University of Science and Technology, Anshan, Liaoning 114051, China (e-mail:2985660842@qq.com).

Shaochuan Xu is a Professor of School of Control Science and Engineering, University of Science and Technology Liaoning, Anshan, 114051 China. (Corresponding author, phone: 86-0412-5929747; e-mail: shaochuanxu1@163.com).

Minghao Ma is a Postgraduate Student of School of Electronic Information, University of Science and Technology Liaoning, Anshan, 114051 China. (e-mail: mmh13591920408@163.com).

significantly reduced model size and improved detection speed by replacing U-Net encoder with MobileNetV3 [8]. Addressing image variations caused by UAV flight altitude fluctuations, L Jianing et al. designed an optimization scheme incorporating EfficientNetV2 [9], coordinate attention mechanisms [10], and CB-Focal Loss functions to enhance model robustness and generalization capability [11].

Although these studies have achieved varying degrees of success in terms of model simplification and performance enhancement, there are still complex challenges in real crusher operating environments: blurring of boundaries between the equipment and the background, varying sizes of ore particles with motion blur, and unpredictable lighting conditions. These factors pose additional difficulties for image recognition and may limit the applicability of existing solutions in specific scenarios. Therefore, how to effectively address these challenges and further improve the accuracy and reliability of image segmentation techniques remains an issue that deserves in-depth discussion.

II. Related Work

A. Crusher Working Process

The actual ore crushing operation takes place in an open air environment. As shown in Figure 1, large haulage vehicles first transport the mined ore to the crusher. Only after the crusher has completed its current task does the next haulage vehicle unload the ore into the crusher. Inside the crusher, the ore is crushed into smaller particles and transported via a conveyor belt to the subsequent processing equipment. During this process, the focus of monitoring is centred on the crusher's inlet, where the system needs to assess its operational status, detect the presence of a complete blockage, and trigger an alarm in time to notify the relevant personnel to clear the material.

In order to ensure the efficiency and safety of the whole operation process, there are several key points that need to be noted and optimised:

1. Transportation and Unloading Coordination: Since the crusher can only process a certain amount of ore at a time, the arrival time and unloading sequence of transport vehicles need to be accurately controlled to avoid production interruptions or build-up problems due to too early or too late unloading.

2. Crusher Condition Monitoring: In addition to basic clogging detection, there is a need for real-time monitoring of the crusher's operating conditions including, but not limited to, motor loads, temperature changes, and vibration. This data can help predict potential failures, take maintenance measures in advance, and reduce unplanned downtime.

3. Intelligent Early Warning System: It is critical to establish an intelligent early warning system based on image recognition and sensor data. The system should not only be able to identify whether the feed opening is clogged or not, but also be able to analyse the degree of clogging and provide appropriate solution suggestions, such as automatically adjusting the feed rate in case of minor clogging, and immediately stopping the feed and alarming in case of serious clogging.

4. Environmental Adaptability: Considering the complexity and variability of the open-air operating environment, including the impact of weather conditions (e.g., rain, snow, wind, sand, etc.) on the operation, the system design needs to give full consideration to its environmental adaptability to ensure that it can still be operated stably under a variety of harsh conditions.

5. Automation and Efficiency Enhancement: The introduction of automation technology can significantly improve production efficiency. For example, the use of automated control systems to optimize the ore conveying speed, according to the actual load of the crusher to dynamically adjust the feed rate, thus maximizing the utilization rate of the equipment while reducing energy consumption.

Through the implementation of the above measures, not only can improve the safety and efficiency of ore crushing operations, but also effectively extend the service life of the equipment and reduce operating costs. This is an important economic value and social benefits for mining enterprises.

B. Detection Challenges

During unloading, clogging may occur due to the collision of large ore particles with the inlet. These blockages fall into two categories: temporary blockages resolved by subsequent ore flow under gravity, and persistent blockages requiring mechanical intervention, as illustrated in Figure 2. Since mechanical clearance is time-consuming and halts operations, accurate detection of persistent blockages is critical for issuing valid alerts.

This study focuses on real-time monitoring of the crusher's inlet status, including determining whether material flow can continue or if blockage alerts must be triggered. While conventional approaches detect ore particles (e.g., Luo et al. [12] applied Mask R-CNN for ore detection and blockage identification in hopper feed ports), such methods face limitations in this scenario:

Ore Detection Challenges:Significant variations in ore particle sizes, coupled with overlapping and dynamic motion (e.g., tumbling), complicate accurate detection. Transient positional changes may cause ore particles to be misclassified as background.Ore detection alone cannot reliably determine feed port status, especially during ambiguous blockage scenarios.

Alternative Approach: Feed Port Detection.Direct detection of the feed port region presents feasibility but introduces new challenges:

(1) Target-Background Similarity: Ore textures and colors may closely resemble the feed port structure, blurring boundaries when particles align with the port.

(2) Lighting Variability and Nighttime Constraints: Outdoor operations under 24/7 working conditions face extreme illumination changes (e.g., overexposure, underexposure), which degrade feature visibility and exacerbate target-background confusion.

(3) Dataset Limitations: Real-world scenarios cannot encompass all edge cases, leading to unreliable model predictions for unobserved conditions.

(4) Detection Speed Requirements: The inlet's state exhibits rapid cyclical changes during continuous ore flow. Slow detection systems risk misinterpreting normal dynamic variations as blockages, causing unnecessary shutdowns or overlooking genuine risks.

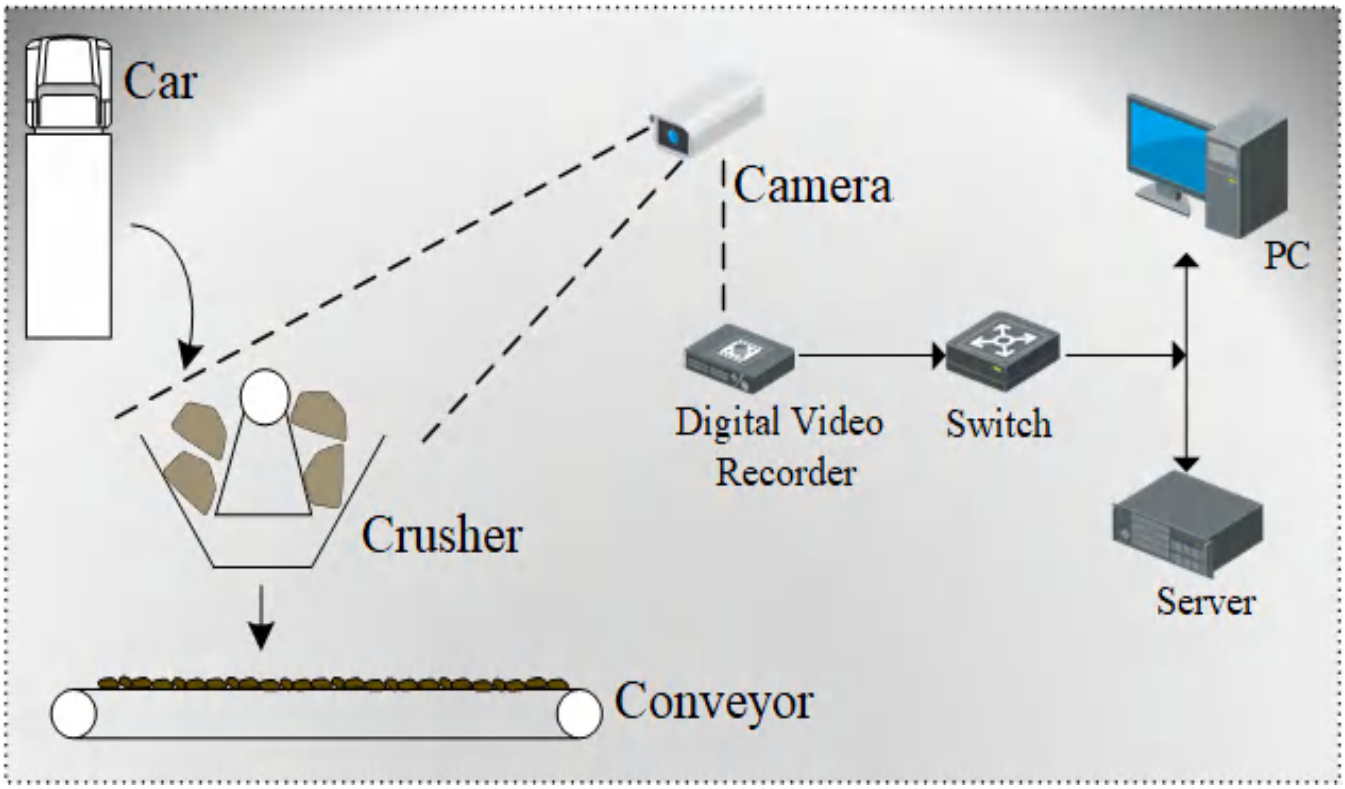


Fig. 1: Crushing Process Flowchart

These factors collectively complicate accurate detection, often resulting in false target-background distinctions.To address these challenges, we propose a multi-scale dilated attention mechanism-enhanced U-Net algorithm for direct feed port detection. Key innovations include:

Data Augmentation: Mitigates lighting variability and dataset limitations through enhanced training diversity. Multi-Scale Dilated Attention [13]: Captures broader contextual features to resolve target-background ambiguity, supplemented by channel attention for feature guidance. Wavelet-based Downsampling: Mitigates texture and boundary feature loss caused by downsampling through wavelet transform, which preserves multi-scale frequency components.

This approach aims to achieve real-time, robust detection under complex industrial conditions.

III. Improved Module

In crusher inlet status detection, two critical factors dominate performance: image preprocessing and model architecture. For dataset optimization, we employ conventional image enhancement techniques to mitigate overexposure and underexposure effects. For algorithmic design, we adopt U-Net as the baseline model and implement targeted improvements.

Fig. 2: Pictures of the crusher scene

A. Image Enhancement

Images captured under natural conditions often suffer from quality degradation due to variable lighting, complex backgrounds, and environmental interference. Common issues include blurring, localized overexposure, and loss of edge details [14]. As illustrated in Figure 3, structural obstructions near the crusher exacerbate lighting disparities during afternoon operations. These artifacts not only degrade visual clarity but also hinder feature extraction, making it challenging for semantic segmentation models to learn target characteristics accurately. Thus, effective preprocessing of raw images is essential before model training to ensure

high-quality data inputs.

By implementing a systematic image preprocessing pipeline, we significantly enhance the quality of crusher inlet images, laying a robust foundation for subsequent deep learning model training. These preprocessing steps improve both training efficiency and model generalization, enabling stronger robustness in complex, real-world operational scenarios.

Fig. 3: Original pictures as well as images with image enhancements

B. Overall Network Architecture

Current image segmentation networks primarily fall into two categories: CNN-based architectures and self-attention mechanism-based networks (e.g., Transformers). Hybrid approaches combining CNN and Transformer layers exist but require intricate architectural tuning. While Transformers often achieve higher accuracy, their computational demands are prohibitive for real-time industrial applications. In contrast, CNN-based models strike a favorable balance between performance and efficiency, making them more deployable on edge devices and preferable for engineering applications.

Our proposed architecture builds upon the classic CNN-based segmentation model U-Net, leveraging its proven effectiveness while addressing domain-specific challenges through targeted modifications.

C. U-Net Adaptation and Proposed Improvements

The original U-Net architecture, designed for medical image segmentation, features a simple encoder-decoder structure with skip connections. The encoder comprises convolutional layers and downsampling operations (initially using max pooling), while the decoder employs upsampling to restore image resolution. A key innovation of U-Net lies in its skip connections between encoder and decoder layers, which mitigate information loss caused by downsampling.

However, segmentation tasks for crusher inlet images face unique challenges: blurred target-background boundaries, large-scale scene complexity, and dynamic lighting conditions. To address these issues, we propose a multi-scale attention-enhanced semantic segmentation model based on U-Net. Proposed Network Architecture

As illustrated in Figure 4, our architecture retains the encoder-decoder framework: Encoder: Utilizes ResNet50 [15] as the backbone for feature extraction from inlet images.

Decoder: Integrates two attention mechanisms—MDCA (Multi-Scale Dilated Convolutional Attention) and SE (Squeeze-and-Excitation) [16] to enhance channel-wise feature understanding and guide global feature extraction.Downsampling: Replaces conventional convolution/pooling with Haar Wavelet Downsampling (HWD) [17], preserving high-frequency details during resolution reduction.

This improvement not only enhances the ability to understand complex scenes, but also effectively copes with the challenges due to changes in ambient light, thus increasing the accuracy and reliability of crusher inlet state monitoring. By combining advanced feature extraction techniques and targeted attention mechanisms, this model realizes effective monitoring of problems such as crusher inlet blockage, demonstrating its practical value and technological advancement.

D. MDCA Module

While traditional U-Net excels at local detail processing, it struggles with large-scale target detection [18]. To resolve this, we introduce the MDCA mechanism, which efficiently encodes contextual information through multi-scale dilated convolutions. Compared to spatial attention mechanisms, MDCA better captures global dependencies while maintaining computational efficiency, thereby enhancing subsequent convolutional operations' ability to extract discriminative features.

In practice, MDCA is inserted into the upsampling stage to refine feature integration through weighted outputs (Figure 5). The module comprises two components:

Multi-Branch Dilated Convolutions: Capture multi-scale contextual information using identical kernel sizes with varying dilation rates. Point-wise Convolution (1×1): Models inter-channel relationships within features. The output of this convolution serves as attention weights to recalibrate MDCA input features.

This design enables adaptive focus on both local details and global structures, significantly improving segmentation accuracy under complex industrial conditions.

The MDCA mechanism can be formally expressed as:

$$At = Conv_{1x1}(\sum_{i=0}^{3} Scale_i(D - Conv(F))) \quad (1)$$

$$out = At \otimes F \quad (2)$$

Let F denote the input feature map. At and Out represent the attention map and output, respectively. The operator "⊗"

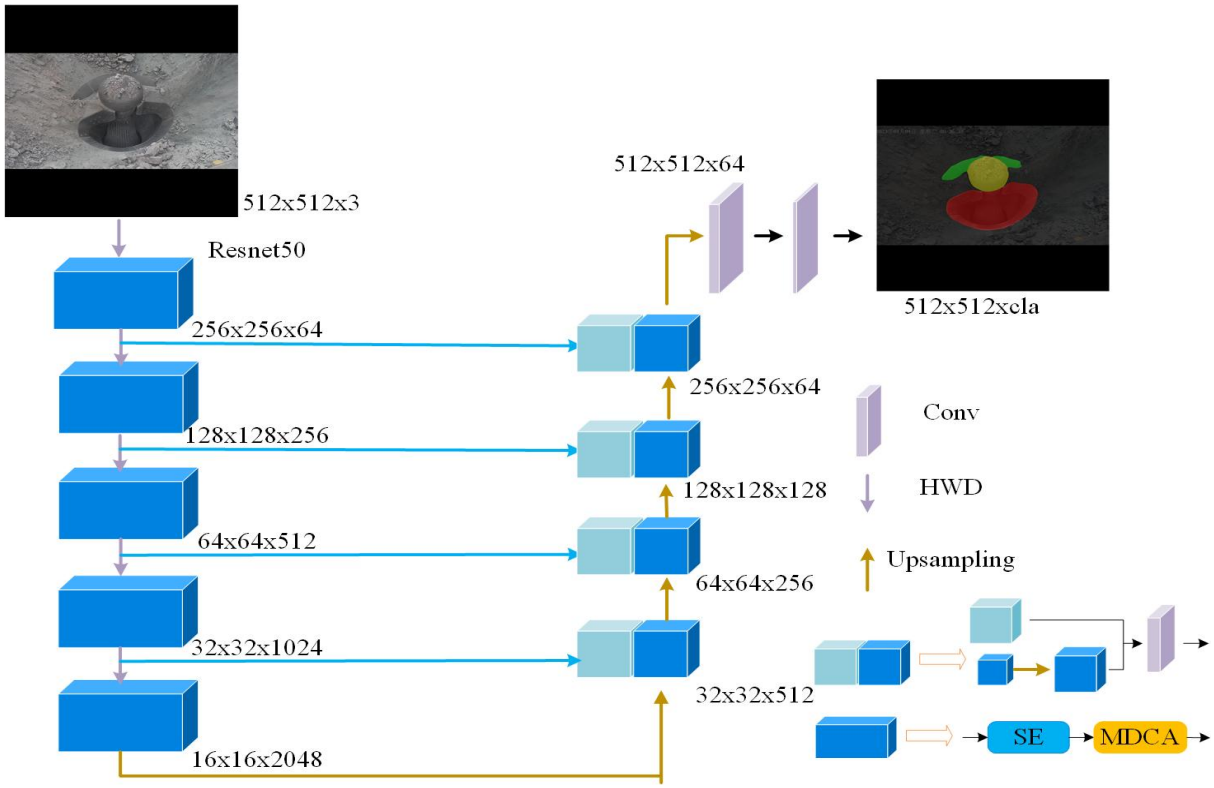


Fig. 4: U-Net model

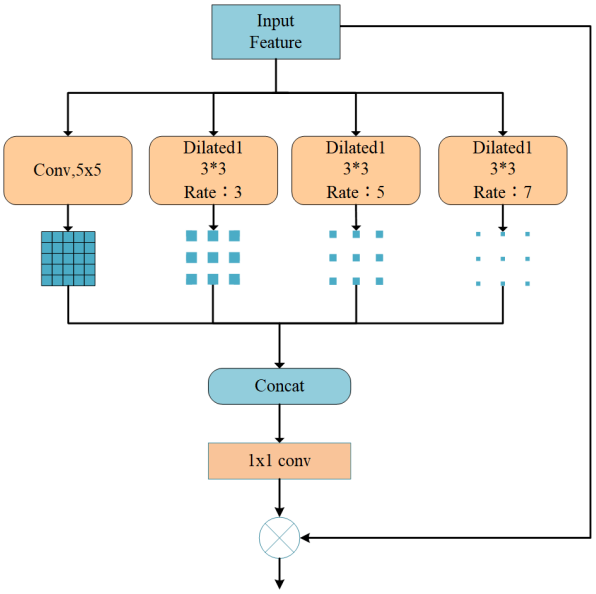


Fig. 5: MDCA

denotes element-wise matrix multiplication. Here, D-Conv refers to dilated convolution, and Scale_i (where i ∈ 0, 1, 2, 3) corresponds to the i-th branch in the architecture, with Scale_0 acting as an identity connection.

Implementation Specifications Kernel Configuration: Each branch employs a kernel size of 3 with dilation rates d = 3, 5, and 7.

Auxiliary Branch: A standard 5×5 convolutional branch is integrated to mitigate checkerboard artifacts caused by dilated convolutions, ensuring feature preservation. Receptive Field Expansion

The effective kernel size R for a dilated convolution with kernel size k and dilation rate d is calculated as:

$$R = k + (k-1) \times (d-1) \tag{3}$$

For dilation rates 3, 5, and 7, the equivalent kernel sizes expand to 7×7, 11×11, and 15×15, respectively. This mechanism achieves multi-scale feature extraction while maintaining parameter efficiency.

E. Haar Wavelet Downsampling (HWD) Module

Outdoor images often suffer from insufficient illumination and uneven lighting conditions, which complicate the extraction of patterns and textures in low-light scenarios[19]. Additionally, the image compression caused by the downsampling process in the encoder further increases the difficulty of feature extraction. To address these challenges, this paper introduces the Haar Wavelet Downsampling (HWD) module. The Haar wavelet transform reduces image resolution without information loss, and signals processed by this transform can be fully reconstructed via inverse transforms, making it widely applicable in image encoding. The core principle involves decomposing images and signals into high-frequency (e.g., edges, textures) and low-frequency (e.g., global structures) components through derivative-based calculations. Since high-frequency components typically contain substantial noise interference, this decomposition effectively reduces noise impact in complex images, thereby improving segmentation accuracy[20].

The HWD module consists of two components (as shown in Figure 6): Feature Encoding:

The input image undergoes a discrete wavelet transform (DWT). Two directional high-frequency feature maps and one low-frequency feature map are selectively retained for subsequent processing. Reconstruction and Downsampling:

The four-channel features obtained from the Haar wavelet transform are concatenated. The concatenated features pass through a sequence of operations: convolutional layers → batch normalization → activation functions → downsampling, ultimately generating the final output.

$$\begin{cases} A(i,j) = \frac{1}{2}\left(I(i,2j) + I(i,2j+1)\right) \\ H(i,j) = \frac{1}{2}\left(I(i,2j) - I(i,2j+1)\right) \\ V(i,j) = \frac{1}{2}\left(I(2i,j) + I(2i+1,j)\right) \\ D(i,j) = \frac{1}{2}\left(I(2i,j) - I(2i+1,j)\right) \end{cases} \tag{4}$$

This design ensures efficient resolution reduction while preserving critical texture details and suppressing noise interference, significantly enhancing segmentation robustness in challenging environments.

F. Loss Function

For the segmentation of crusher inlet images under natural conditions, this study adopts a combined loss function integrating Cross-Entropy (CE) loss and Dice loss[21]. The CE loss function guides the model to learn pixel-level class probability distributions, improving per-pixel classification accuracy, denoted as CE. The Dice loss addresses boundary ambiguity issues, denoted as Dice[22]. The combination of these two loss functions balances global and local segmentation performance. Formulation:

$$\mathcal{L}_{\text{CE}} = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \log(\hat{y}_i) + (1-y_i)\log(1-\hat{y}_i)\right] \tag{5}$$

$$\mathcal{L}_{\text{Dice}} = 1 - \frac{2\sum_{i=1}^{N} y_i \hat{y}_i}{\sum_{i=1}^{N} y_i + \sum_{i=1}^{N} \hat{y}_i} \tag{6}$$

$$\mathcal{L} = L_{\text{CE}} + L_{\text{Dice}} \tag{7}$$

Here, λ is a weighting parameter that balances the contributions of the two losses[23]. This hybrid approach enhances segmentation accuracy for both fine-grained details and ambiguous boundaries, particularly under complex lighting and texture conditions.

IV. Experiments

In this section, we provide a comprehensive evaluation of the performance of the improved U-Net in the crusher inlet segmentation task.

First, we verify the effectiveness of the proposed module through a series of ablation experiments. These experiments aim to clarify the specific contributions of each component to the overall performance. Additionally, we conducted comparison experiments between the approach proposed in this paper and several of the most mainstream semantic segmentation algorithms currently available. The experimental results demonstrate that our method outperforms other algorithms in key performance metrics such as MIoU (mean intersection over union) and mPA (mean pixel accuracy), thereby highlighting its significant advantages in terms of image segmentation accuracy and real-time processing capability in complex environments[24].

To more intuitively demonstrate the performance of the improved U-Net, we visualized its segmentation results and compared them with those of existing models. This visual analysis not only clearly highlights the advantages of the improved model but also provides valuable references for further optimization.

Through the evaluation and analysis of the above system, we verified the efficiency and reliability of the improved U-Net in practical applications, laying a solid foundation for future research and practice.

A. Dataset and Implementation

The experimental dataset was constructed from surveillance videos (resolution: 2560×1440) of operational crushers at a Chinese iron ore mining site. Images were extracted at 2-second intervals and manually annotated after rigorous filtering to exclude scenarios with rain, snow, or heavy dust. The final dataset comprised 300 images, partitioned into training (70%), validation (20%), and test (10%) sets. Temporal diversity was ensured by categorizing data into seven timeframes (midnight, dawn, morning, noon, afternoon, dusk, night), with 10 unblocked inlet images selected per timeframe. Additionally, 160 operational state images depicting partial blockages were included. Professional annotators labeled three critical regions: Port A (lower feed port), Port B (upper feed port), and beam cap (structural support), generating pixel-accurate segmentation masks.

The experiments were conducted on a Windows 11 system equipped with an NVIDIA RTX 4060 GPU (8GB VRAM), 16GB RAM, and 30GB virtual memory. The software stack utilized Python 3.8.0 and PyTorch 2.2.1. Model training employed a hybrid Dice-Cross-Entropy loss function to address class imbalance, optimized via Adam (momentum=0.9) with an initial learning rate of 0.001 and cosine annealing scheduling. Training spanned 80 epochs with a batch size of 8, balancing computational efficiency and convergence stability.

B. Evaluation Metrics

Three metrics were adopted:

Mean Pixel Accuracy (mPA): Computed as the average ratio of correctly classified pixels to total pixels per image. This metric evaluates the overall accuracy of pixel-wise classification on a per-image basis and then averages these accuracies across all images.

Mean Intersection over Union (mIoU): Quantifies the overlap between predicted and ground-truth regions. For each class, IoU is calculated as the intersection of predicted and ground-truth regions divided by their union.

Pixel Accuracy (Acc): Similar to mPA but often used to refer to the global accuracy across the entire dataset rather than averaged per image. It gives the overall percentage of correctly classified pixels in the entire dataset:

$$PA = \frac{TP + TN}{TP + TN + FP + FN} \tag{8}$$

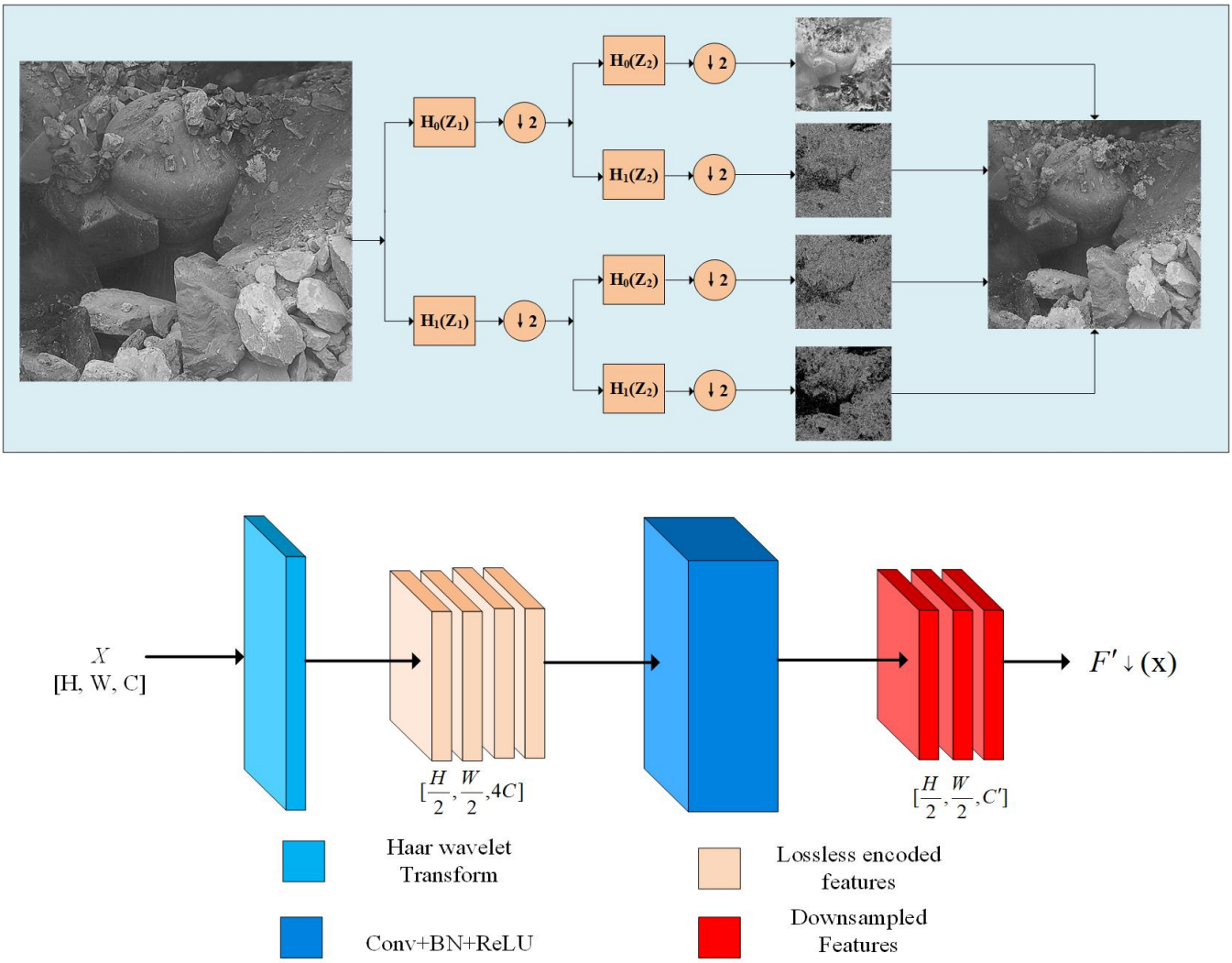


Fig. 6: HWD

$$mPA = \frac{1}{n}\sum_{i=1}^{N} PA_i \tag{9}$$

$$IoU = \frac{TP}{TP + FP + FN} \tag{10}$$

$$mIoU = \frac{1}{n}\sum_{i=1}^{N} IoU_i \tag{11}$$

These metrics provide a comprehensive evaluation of the segmentation model's performance, covering both per-image and global accuracy, as well as the overlap quality between predicted and ground-truth regions.

C. Ablation Study

The proposed network architecture modifies the backbone structure while integrating attention mechanisms and Haar wavelet downsampling. To evaluate the effectiveness of each component, ablation experiments were conducted by incrementally removing modules from the improved UNet framework.

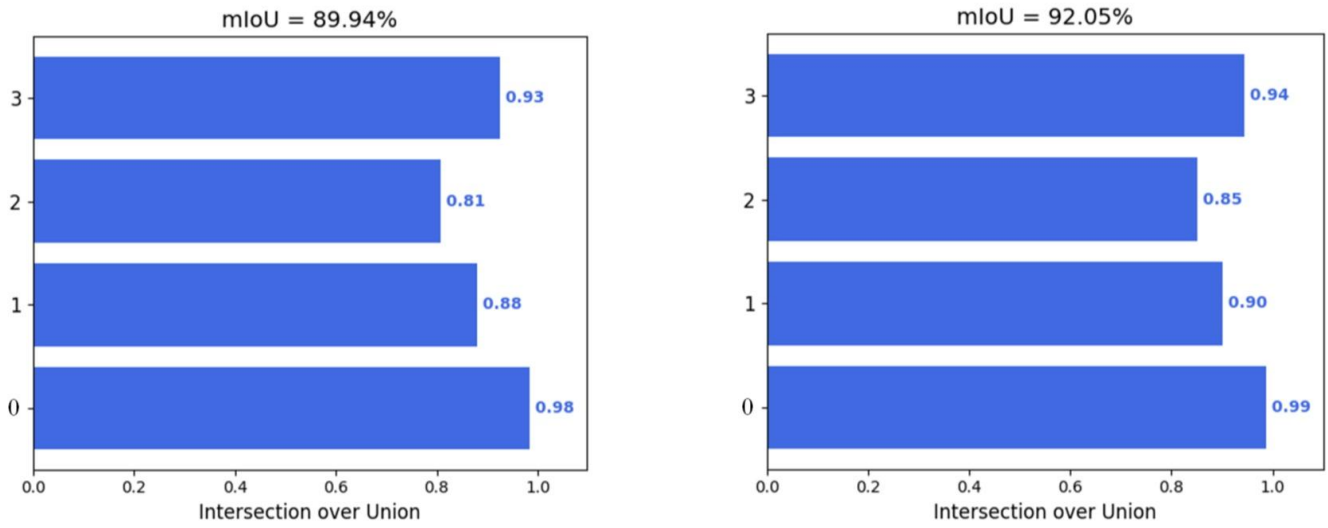
Table 1 presents the ablation results on the dataset, where checkmarks (✓) denote the inclusion of specific modules. All experiments were conducted using data augmented with identical image enhancement techniques. Here, "SE" and "SE + MDCA" represent two attention mechanism configurations.

TABLE I
RESULTS OF ABLATION STUDY PERFORMED

Method	Attention		HWD	mPA(%)	Miou(%)	Acc(%)
	SE	MDCA+SE				
Base	–	–	–	89.12	94.50	96.80
1	✓	–	–	90.04	95.27	97.59
2	–	✓	–	90.09	94.71	98.56
3	✓	–	✓	91.27	95.35	98.77
4	–	✓	✓	92.05	96.29	98.87

Key observations include: 1. Attention Mechanisms: Both SE and MDCA modules achieve modest improvements in detection accuracy. Their primary contribution lies in mitigating background-target misclassification errors. Due to MDCA's higher computational overhead, it is selectively applied to critical regions.

2. Haar Wavelet Downsampling: This module marginally improves performance (details omitted in original text), suggesting that while its feature filtering aids detection, its impact is less pronounced compared to attention mechanisms.



(a) U-Net (b) ours

Fig. 7: mIoU for different models. (a) .Original U-Net model in the test set of the current dataset for each category of IoU situation. (b) . Our model in the test set of the current dataset for each category of IoU situation

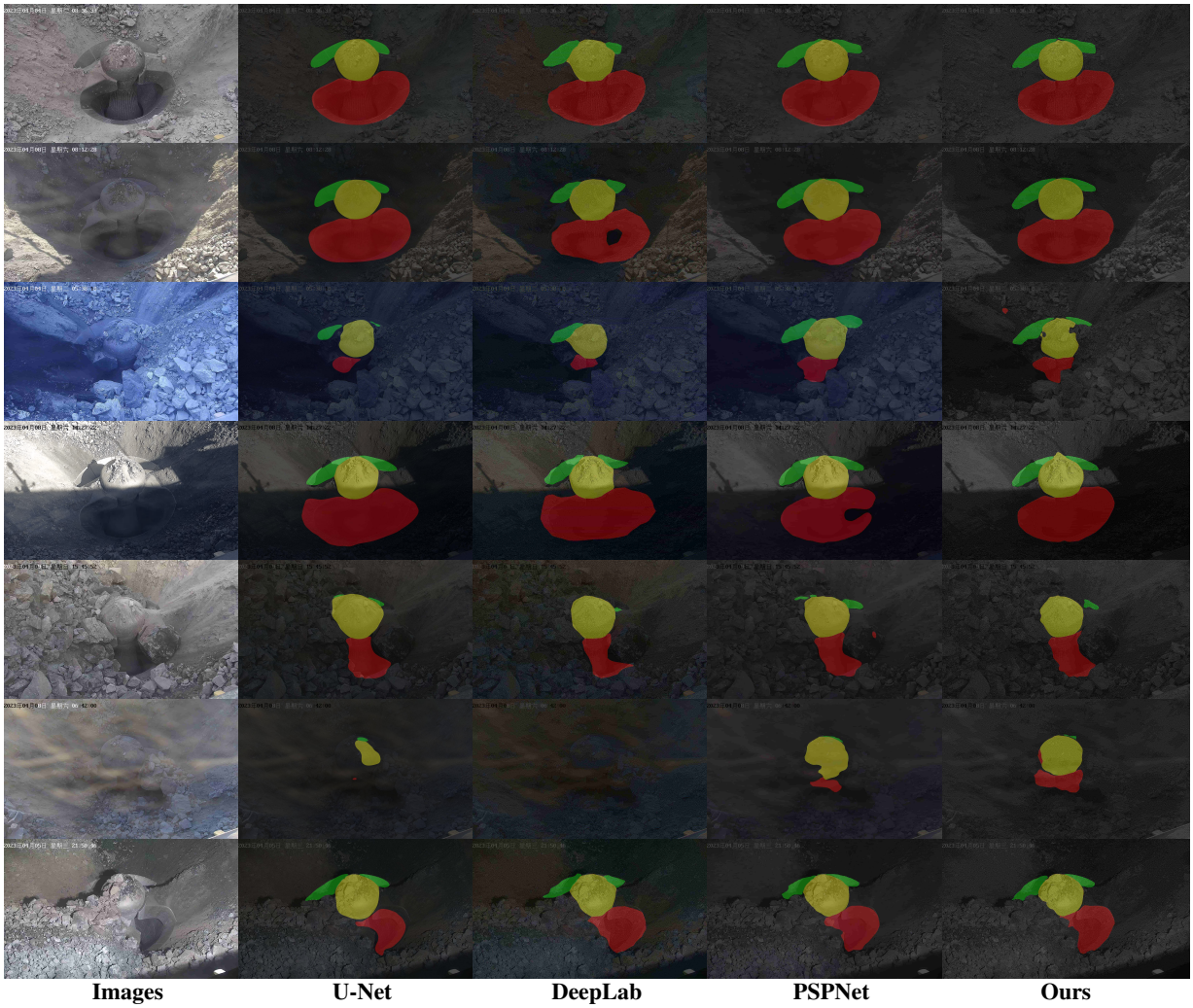


Fig. 8: Instantiated test results

Critically, the synergistic integration of all modules achieves an optimal balance between performance and efficiency. The proposed algorithm greatly improves the segmentation accuracy, where mIoU (92.05%) and mPA (96.29%) meet the strict practical criteria.

D. Comparison with Mainstream Models

To rigorously validate the superiority of the proposed framework, comprehensive comparative experiments were conducted against multiple state-of-the-art models. Each model configuration underwent five independent trials under identical experimental conditions, including the same training dataset, hyperparameters, and hardware environment. Performance metrics—mPA, mIoU, parameter count (Params), and Acc.

For equitable comparison, all baseline models (original UNet, DeepLab v3+, and PSPNet) adopted ResNet50 as their backbone architecture.

TABLE II
COMPARISON OF DIFFERENT MODELS

Module	mPA(%)	Miou(%)	Acc(%)
Unet	89.12	94.50	97.59
PSPNet	87.81	94.46	98.35
Deeplab	92.01	96.01	98.83
Ours	92.05	96.29	98.87

Table 2 summarises the results of comparing the improved U-Net model with the three classical semantic segmentation models. Figure 7 show that our model performs more accurately in the crusher inlet segmentation task, significantly outperforming the three classical models in terms of mean pixel accuracy (mPA) and mean intersection and merger ratio (mIoU). Despite the increase in the number of parameters, we succeeded in minimising the performance impact of the increase in the number of parameters by applying quantisation and cropping techniques in all deployment environments.

In addition, it can be more clearly observed from the visualised comparison graphs that the model proposed in this paper shows more stable and consistent detection results, further validating its superiority.

Figure 8 shows the visualised detection results of each group of models under different lighting conditions and crusher operating conditions. It is observed that all models can obtain good detection results when the light is weak but evenly distributed. However, when entering the afternoon hours, with significant changes in the light, PSPNet and DeepLabV3+ showed obvious voids in detecting the no-feed state in the lower half of the crusher, and it was difficult to accurately identify the boundary of the crusher inlet. In contrast, the method proposed in this paper still performs well in this situation without similar problems. In addition, in the third set of images, the refractive effect of sunlight at sunrise causes the crusher image to exhibit a bluish hue, which makes it challenging for all models, including the method in this paper, to achieve the desired detection results. Nevertheless, the method in this paper shows higher robustness and accuracy compared to other models, especially in dealing with complex lighting conditions. This comparison shows that the improved method proposed in this paper has obvious advantages in dealing with variable lighting conditions and object detection tasks in complex environments.

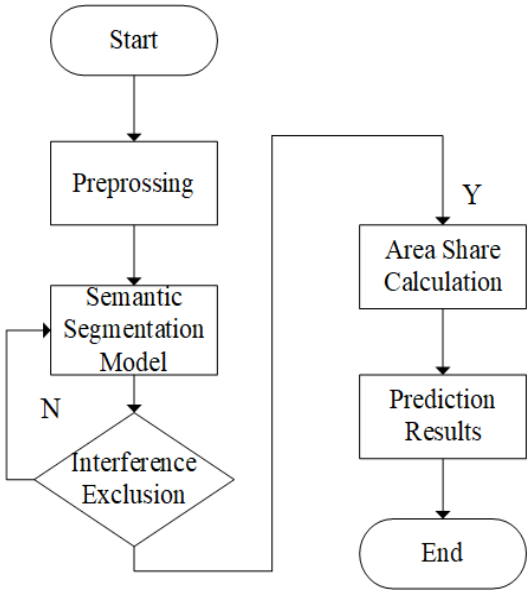


Fig. 9: System Flowchart

V. System Overview

A. System Architecture

The proposed system employs a camera for image acquisition, utilizing the OpenCV library to decode video streams, with a sampling rate of two frames per second for semantic segmentation tasks. The workflow is structured as follows:

1. Preprocessing: Raw images undergo preliminary enhancement via traditional machine vision techniques before being fed into the semantic segmentation model.

2. Baseline Calibration: A no-load reference image (captured during system initialization) establishes the baseline pixel count for subsequent state comparisons.

3. State Inference: Crusher discharge status is determined by analyzing pixel quantity ratios relative to the no-load baseline and monitoring temporal state transitions. Beam cap detection identifies severe environmental interference, triggering a hysteresis-based response mechanism to ensure operational stability.

4. Human-Machine Interface (HMI): A LabVIEW-based interface visualizes real-time crusher status, segmentation results, and critical alerts, enabling rapid operator intervention.

B. Human-Machine Interface

Figure 8 shows the HMI system integrates the following functional modules:

1. Login interface: User authentication through password-protected access.

2. Detection interface: Real-time pixel area profile of port A/B (x-axis: time; y-axis: number of target pixels). Dual-port status logic: Normal: Both ports are within baseline thresholds. Warning: Single port deviation triggers

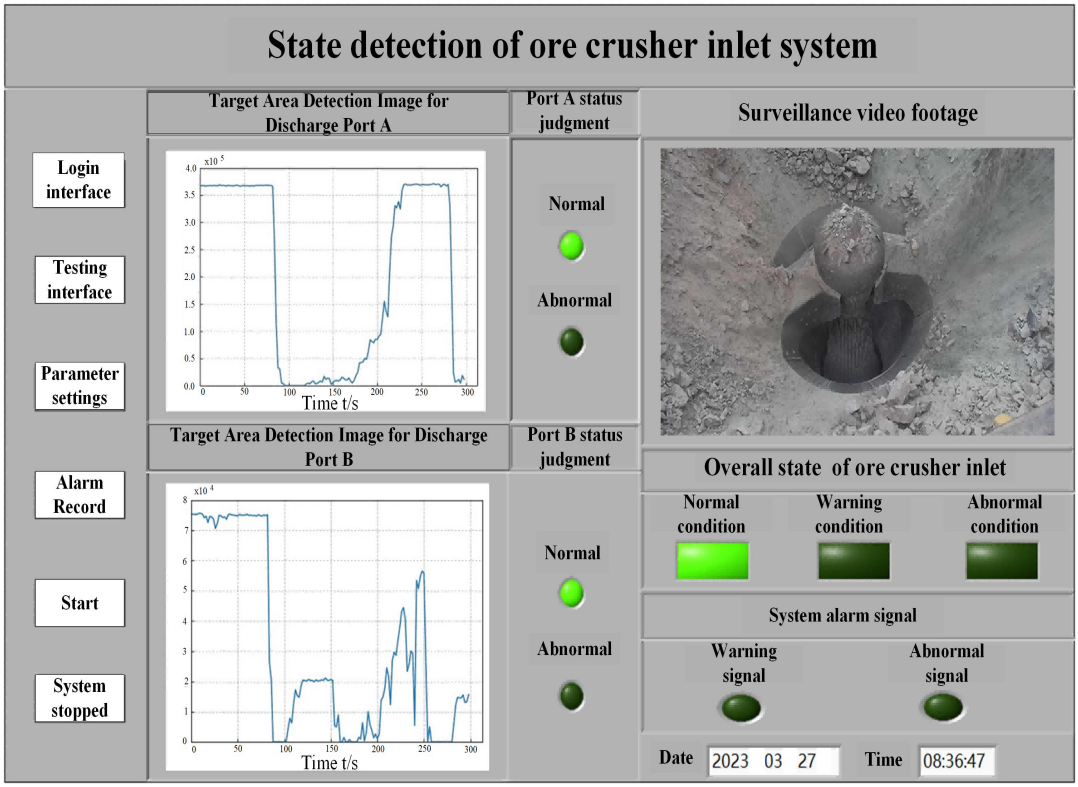


Fig. 10: System

an alarm. Critical: Dual port anomaly triggers an emergency alarm. All statuses are expressed in the form of indicator lights, and when the light is on, it is in the corresponding state. 3. Real-time monitoring:

3.Real-time display of monitoring and control. real-time monitoring: real-time display of the monitoring site screen.

4. Configuration panel: Adjustable alarm thresholds and hysteresis parameters.

5. Alarm Record: Time-stamped record of abnormal operation.

6. start operation and system stop: control the detection system to start normal operation, as well as the system stop.

VI. Conclusion

This study focuses on crusher inlet status detection by improving the U-Net algorithm and evaluating its performance on real-world data collected from open-pit crushers. The proposed method targets fixed-location scenarios where insufficient contrast between the environment and the crusher occurs during certain periods due to lighting conditions. By integrating image enhancement with MDCA+SE attention mechanisms, the issue of erroneous detection of the crusher inlet location is effectively resolved. Haar wavelet downsampling is adopted to retain more texture and edge features typically lost during downsampling.

Experimental results demonstrate that the improved model achieves satisfactory accuracy and speed in practical crusher inlet status detection tasks, validating the method's applicability. However, the current study is limited to specific operational sites, and the model lacks generalization capability for diverse industrial environments.

References

[1] Ronneberger, Olaf, Philipp Fischer, and Thomas Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, Munich, Germany, 2015, pp. 234–241.

[2] Zhao, Hengshuang, et al., "Pyramid scene parsing network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, HI, USA, 2017, pp. 2881–2890.

[3] Chen, Liang-Chieh, et al., "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proceedings of the European Conference on Computer Vision*, Munich, Germany, 2018, pp. 833–851.

[4] Gao Changxin, et al., "Deep learning-based real-time semantic segmentation: A survey," *Journal of Image and Graphics*, vol. 29, no. 5, pp. 1119–1145, 2024.

[5] Chicchon, Miguel, et al., "Semantic segmentation of fish and underwater environments using deep convolutional neural networks and learned active contours," *IEEE Access*, vol. 11, pp. 33652–33665, 2023.

[6] Wang, Wei, et al, "An improved boundary-aware U-Net for ore image semantic segmentation," *Sensors*, vol. 21, no. 8, p. 2615, 2021.

[7] Liu, Jingyi, Hanquan Zhang, and Dong Xiao, "Research on anti-clogging of ore conveyor belt with static image based on improved Fast-SCNN and U-Net," *Scientific Reports*, vol. 13, no. 1, p. 17880, 2023.

[8] Jin, Ge, et al, "An end-to-end steel surface classification approach based on EDCGAN and MobileNet V2," *Sensors*, vol. 23, no. 4, p. 1953, 2023.

[9] Tan, Mingxing, and Quoc Le, "EfficientNetV2: Smaller models and faster training," in *Proceedings of the International Conference on Machine Learning*, PMLR, 2021, pp. 10096–10106.

[10] Bastidas, Alexei A., and Hanlin Tang, "Channel attention networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, Long Beach, CA, USA, 16–17, 2019, pp. 881-888.

[11] Jianing L, Zhao Z, Xiaohang et al., "Wheat lodging types detection based on UAV image using improved EfficientNetV2," *Smart Agriculture*, vol. 5, no. 3, pp. 62–73, 2023.

[12] Howard, Andrew, et al., "Searching for MobileNetV3," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, Seoul, South Korea, 2019, pp. 1314–1324.

[13] Luo Xiaoyan, Liu Shun, Wang Xingwei et al., "Research on identification and location of blocked ore at ore bin inlet based on Mask RCNN," *Nonferrous Metals Science and Engineering*, vol. 13, no. 1, pp. 101–107, 2022.
[14] Wenyu Liu , Wentong Li and Jianke Zhu et al., "Improving nighttime driving-scene segmentation via dual image-adaptive learnable filters," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 33, no. 10, pp. 5855–5867, 2022.
[15] Guo, Meng-Hao et al., "SegNeXt: Rethinking convolutional attention design for semantic segmentation," *arXiv preprint arXiv:2209.08575*, 2022. [Online]. Available: https://doi.org/10.48550/arXiv.2209.08575
[16] N. Ma, X. Zhang, H. T. Zheng, and J. Sun, "ShuffleNet V2: Practical guidelines for efficient CNN architecture design," in *Computer Vision – ECCV 2018*, vol. 11218, pp. 116–131, 2018.
[17] G. Xu, W. Liao, X. Zhang et al., "Haar wavelet downsampling: A simple but effective downsampling module for semantic segmentation," *Pattern Recognition*, vol. 143, p. 109819, 2023.
[18] E. Batziou, K. Ioannidis, I. Patras et al., "Low-light image enhancement based on U-Net and Haar wavelet pooling," in *Proceedings of the International Conference on Multimedia Modeling*, Lisbon, Portugal, 6–10, 2023, pp. 510–522.
[19] Yuan Luo, Jixiang Shen, and Fangyu Li, "Real-time Visual SLAM Based on Lightweight PSPNet Network," *Engineering Letters*, vol. 32, no. 10, pp. 1981-1992, 2024.
[20] Huikai Li, and Jie Wu, "LSOD-YOLOv8s: A Lightweight Small Object Detection Model Based on YOLOv8 for UAV Aerial Images," *Engineering Letters*, vol. 32, no. 11, pp. 2073-2082, 2024.
[21] Haitian Qin, and Yang Xu, "Occluded Pedestrian Re-identification Method Based on Multi-scale Feature Fusion," *Engineering Letters*, vol. 32, no. 12, pp. 2378-2390, 2024.
[22] Shao, X., Lyu, Z., Li, H. et al., "Denet: An Effective and Lightweight Real-time Semantic Segmentation Network for Coal Flow Monitoring," *Journal of Real-Time Image Processing*, vol. 22, no. 1, pp. 1-16, 2025.
[23] Li X, Li S, Dong L, et al., "An Image Segmentation Method of Pulverized Coal for Particle Size Analysis," *International Journal of Mining Science and Technology*, vol. 33, no. 9, pp. 1181-1192, 2023.
[24] I. Bello, "LambdaNetworks: Modeling Long-Range Interactions Without Attention," *arXiv preprint arXiv:2102.08602*, 2021. [Online]. Available: https://arxiv.org/abs/2102.08602