Frontiers in Computing and Intelligent Systems
ISSN: 2832-6024 | Vol. 14, No. 3, 2025

A Non-contact Real-time Diameter Measurement Method for Steel Wire Ropes via W-EDSR and W-DeblurNet

Hang Shen [1], Jiafeng Ren [1], Kuosheng Jiang [2,*], Yuanyuan Zhou [2]

[1] The State Key Laboratory of Digital and Intelligent Technology for Unmanned Coal Mining, Anhui University of Science and Technology, Huainan Anhui, 232000, China
[2] The School of Mechatronics Engineering, Anhui University of Science and Technology, Huainan Anhui, 232000, China

*** Corresponding author:** Kuosheng Jiang

Abstract: Steel wire ropes are widely used in hoisting, transportation, mining and marine engineering, and their complex braided structure provides excellent tensile strength and fatigue resistance. To overcome the limitations of conventional methods for detecting steel wire rope diameters—particularly in terms of accuracy, robustness, and adaptability to complex working environments—this study proposes a non-contact, real-time detection approach based on super-resolution (SR) technology. The method employs an RGB-Depth camera to simultaneously acquire both visual and depth data of the steel wire rope, and leverages an enhanced super-resolution network (W-EDSR) in combination with a self-supervised deblurring network (SRN-WDeblur) to enhance image clarity and mitigate motion blur. Following this, a semantic segmentation and edge extraction algorithm, augmented with depth information, is applied to derive continuous and smooth rope boundaries. Experimental results indicate that this method achieves 99.35% accuracy for static steel wire rope measurements and maintains 99.24% accuracy when the rope moves at 10 m/s, significantly outperforming traditional approaches and demonstrating strong robustness against motion blur. This research provides a reliable and efficient solution for online monitoring of mine hoisting steel wire ropes, contributing to enhanced intelligence and safety in mining operations.

Keywords: Steel Wire Rope; Super-resolution; Machine Vision; Deep Learning.

1. Introduction

Image quality is critical for visual measurement of steel wire rope diameter, yet high-speed motion, vibration, poor lighting, and dust often cause low resolution and motion blur. These issues blur edges and remove texture details, hindering accurate extraction of the two outer tangent points required for diameter calculation. Although existing methods work under good imaging conditions, they struggle with blurred and low-resolution images, are often too computationally heavy for real-time use, and remain sensitive to edge clarity. Therefore, a method that enhances image sharpness efficiently, resists motion blur, and supports real-time deployment is urgently needed for reliable diameter detection under complex working conditions. The contour of the rope image can be approximated as the envelope of a group of delayed sinusoids, where each sine wave represents the kth strand. As shown in Figure 1.

With the development of non-destructive testing methods, vision-based detection technologies have become a research hotspot among many scholars [1], [2].Steel wire ropes, composed of multiple thin steel wires twisted into strands or bundles, serve as flexible cables and play critical roles in lifting, traction, tensioning, and load-bearing in engineering industries such as coal mining, elevators, and bridge cables. However, due to prolonged service periods and repeated tensile forces, steel wire ropes often develop defects such as wear and wire breakage during operation. These defects result in a reduction in rope diameter and degradation of elasticity, which severely impair their load-bearing capacity and reliability. Therefore, accurate and timely detection of steel wire rope diameter, along with reliability evaluation, is of great significance for ensuring the safe operation of mining production.

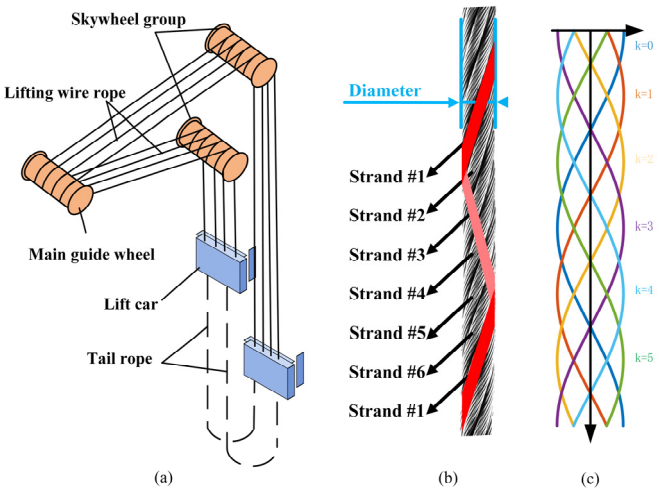

Figure 1. Illustration of the physical model. (a) The mine hoisting system. (b) The structure of the steel wire rope. (c) Approximate image

There are two main methods for measuring the diameter of steel wire ropes: offline measurement and online measurement. Offline measurement is typically conducted manually in a stationary state, using contact-based tools such as vernier calipers, outside micrometers, or specialized gauges to measure the outer diameter of the steel wire rope at different positions. This process requires measuring the distance between the upper and lower outer tangent lines of at least two adjacent strands of the rope. Although this method

provides high measurement accuracy, it demands substantial manpower, time, and effort, and is unsuitable for long-distance ropes or enclosed operating environments. Moreover, it cannot be used for real-time diameter monitoring. In recent years, with the advancement of non-contact sensing technologies, online measurement techniques have become a focal point of research for many scholars [3], [4], [5], [6], [7], [8]. Currently, online non-destructive diameter detection mainly includes laser scanning, inductive measurement, and machine vision methods. The laser scanning method employs high-precision laser distance sensors to emit laser beams and receive the reflected signals from the surface of the rope. Combined with 3D reconstruction algorithms, it accurately calculates the rope's diameter. While this method offers non-contact and high-precision advantages, it is easily affected by rope vibration and surface reflection under operating conditions and imposes strict requirements on sensor installation positions. The inductive measurement method estimates diameter changes based on the influence of the rope on the inductance value as it passes through a sensor coil. It features simple structure and low cost but is sensitive to external factors such as temperature and electromagnetic interference, making it suitable only for rough trend detection. In contrast, the machine vision method captures clear images of the steel wire rope using high-resolution industrial cameras, accurately identifies the rope's contour edges through image processing, and calculates its pixel width, which is then converted to a physical diameter using calibration parameters. However, since image clarity and edge sharpness directly affect measurement accuracy, traditional machine vision methods have high requirements for lighting conditions and background contrast. Therefore, in recent years, an increasing number of researchers have incorporated deep learning models into traditional vision methods to improve robustness and measurement accuracy, especially for industrial online detection tasks under complex backgrounds [9], [10].

Visual inspection technology is a modern measurement method that integrates computer image processing, information analysis, optoelectronics, and pattern recognition. It offers significant advantages such as non-contact operation, high response speed, and low cost, and has been widely applied in the field of automated precision inspection due to its real-time and online capabilities [11], [12]. In recent years, many scholars have conducted extensive research on steel wire rope inspection by combining visual inspection technology with deep learning. Wang et al. [13] used a camera to sample the steel wire rope before underground operations and proposed a YOLOv5-based object detection module to accurately detect external defects of the rope. In addition, the algorithm incorporated transfer learning techniques, which effectively improved the detection accuracy and addressed the challenges of defect detection in mining steel wire ropes. Dong et al. [14] proposed a multi-view scanning system that utilizes novel high-resolution image scanning hardware along with an advanced panoramic image stitching algorithm. Corrosion is accurately identified through deep learning techniques. The system has been validated and applied in the inspection and evaluation of replaced suspended steel wire ropes. Zhou et al. [15] proposed a deep learning-based VPT framework, which integrates an image preprocessing (IP) scheme with a deep convolutional neural network (DCNN). The IP scheme is designed to eliminate background interference and normalize the image data, while the DCNN—an improved version based on LeNet-5—is employed to extract features from datasets under various working conditions. Experimental results demonstrate that the proposed framework can accurately extract the region of interest (ROI) from steel wire rope images and achieves a detection accuracy of 95.55%. Yu et al. [16] proposed a novel deep learning detection network based on a fully convolutional one-stage architecture, termed CABF-FCOS. The framework adopts an anchor-free FCOS as the detection backbone and introduces a Channel Attention Mechanism (CAM) module to reduce the loss of feature information. Additionally, a Bidirectional Feature Fusion Network (BFFN) is employed to replace the conventional Feature Pyramid Network (FPN). Experimental results demonstrate that the proposed method outperforms current state-of-the-art (SOTA) detection approaches in terms of accuracy and meets the real-time detection requirements of industrial applications.

The aforementioned studies demonstrate that combining visual inspection with deep learning is an effective approach for steel wire rope detection [17], [18]. To improve detection accuracy, enhancing the quality of input images is a common strategy in visual inspection. Currently, widely adopted image quality enhancement methods mainly follow two technical paths. The first type consists of traditional image enhancement techniques, such as gamma correction and Contrast Limited Adaptive Histogram Equalization (CLAHE). These methods are simple to implement and offer practical utility, which has led to their widespread adoption. However, they may introduce additional noise during processing, reducing the contrast between target regions and the background, thereby affecting the accuracy of subsequent detection. Thus, appropriate adjustments and optimizations are needed in practical applications to avoid interference with key image information. The second type is based on Super-Resolution (SR) reconstruction techniques, which restore compressed or blurred details in images and convert low-resolution images into higher-resolution versions. SR techniques hold significant value in image processing, computer vision, and reconstruction tasks, as they can effectively enhance image quality and reveal more detailed information [19], [20]. Benefiting from continuous technological advancements, SR has shown strong capabilities in restoring image details, greatly improving visual quality and information richness, and has been widely applied in fields such as remote sensing imagery, medical imaging, and industrial inspection [21], [22].Feng et al. [23] proposed a fast and accurate crack evaluation method that combines multi-sensor fusion SLAM (Simultaneous Localization and Mapping) with image super-resolution technology. This approach enables the direct acquisition of textured point clouds from bridge structures, significantly improving both the efficiency and the accuracy of crack width calculation. Chen et al. [24] proposed a Fourier attention detection network guided by a super-resolution (SR) branch. By introducing an SR branch, the network effectively leverages high-resolution image information to enhance defect detection performance. A series of experiments demonstrated that the proposed method achieves a good balance between detection efficiency and accuracy in patterned wafer defect detection tasks. Wang et al. [25] designed a parallel fusion architecture based on Convolutional Neural Networks (CNN) and Transformers, named FPFCT. The architecture mainly consists of a Flexible Parallel Fusion (FPF) block and a Dynamic Edge Attention (DEA) network, which effectively avoids irreversible

information loss caused by the inherent sampling limitations of CNNs and Transformers. It also dynamically enhances the contour features of multi-scale targets. Experimental results show that FPFCT outperforms most state-of-the-art (SOTA) super-resolution (SR) methods. Sun et al. [26] proposed a super-resolution reconstruction algorithm called Multi-scale Residual Dilated Convolution Attention Network (MRDCAN). This method employs multi-scale dilated convolutions to expand the receptive field of convolutional layers and introduces a loss function that combines Mean Squared Error (MSE) and Structural Similarity Index Measure (SSIM) for image quality evaluation. The algorithm effectively eliminates artifacts and enriches the contour details of wafer images, thereby meeting high image quality requirements. Studies have shown that the proposed algorithm achieves an SSIM of 94.26% on reconstructed super-resolution wafer images.

In this study, the Enhanced Deep Super-Resolution Network (EDSR) is selected as the primary method for image quality enhancement [27]. EDSR is a classical deep learning model that employs residual blocks with batch normalization removed, effectively improving the model's ability to restore image details. Compared with traditional super-resolution methods, EDSR achieves a good balance between image reconstruction quality and computational efficiency, making it particularly suitable for detection tasks in industrial scenarios where high image precision is required.

Previous studies have verified that the quality of input images significantly affects detection accuracy. Therefore, enhancing image quality has become a key step in improving the performance of steel wire rope detection. Building upon the work of earlier researchers, this paper proposes a steel wire rope diameter detection method based on the Enhanced Deep Super-Resolution Network (EDSR), which integrates image deblurring and super-resolution reconstruction techniques. The main steps are as follows:

1) Motion blur suppression is applied to the steel wire rope images using W-DeblurNet, a deblurring network with a scale-progression structure, in order to reduce the impact of speed variation on image quality.

2) Image resolution is further enhanced by employing W-EDSR, a lightweight super-resolution network, which improves edge sharpness and texture detail while optimizing the network structure to meet real-time detection requirements.

3) A theoretical model for steel wire rope diameter detection is established by integrating the image enhancement module with an edge extraction algorithm. The boundaries of the steel wire rope are accurately extracted from high-quality images, and the diameter is calculated using the least squares method.

The structure of this paper is organized as follows. Section 2 is devoted to the description of the proposed method, where the overall framework and implementation details of each component are presented. Experimental validation and performance evaluation based on real-world and simulated data are conducted in Section 3. Finally, the conclusion of this study is drawn in Section 4.

2. The Proposed Method

2.1. Architecture of the Proposed Method

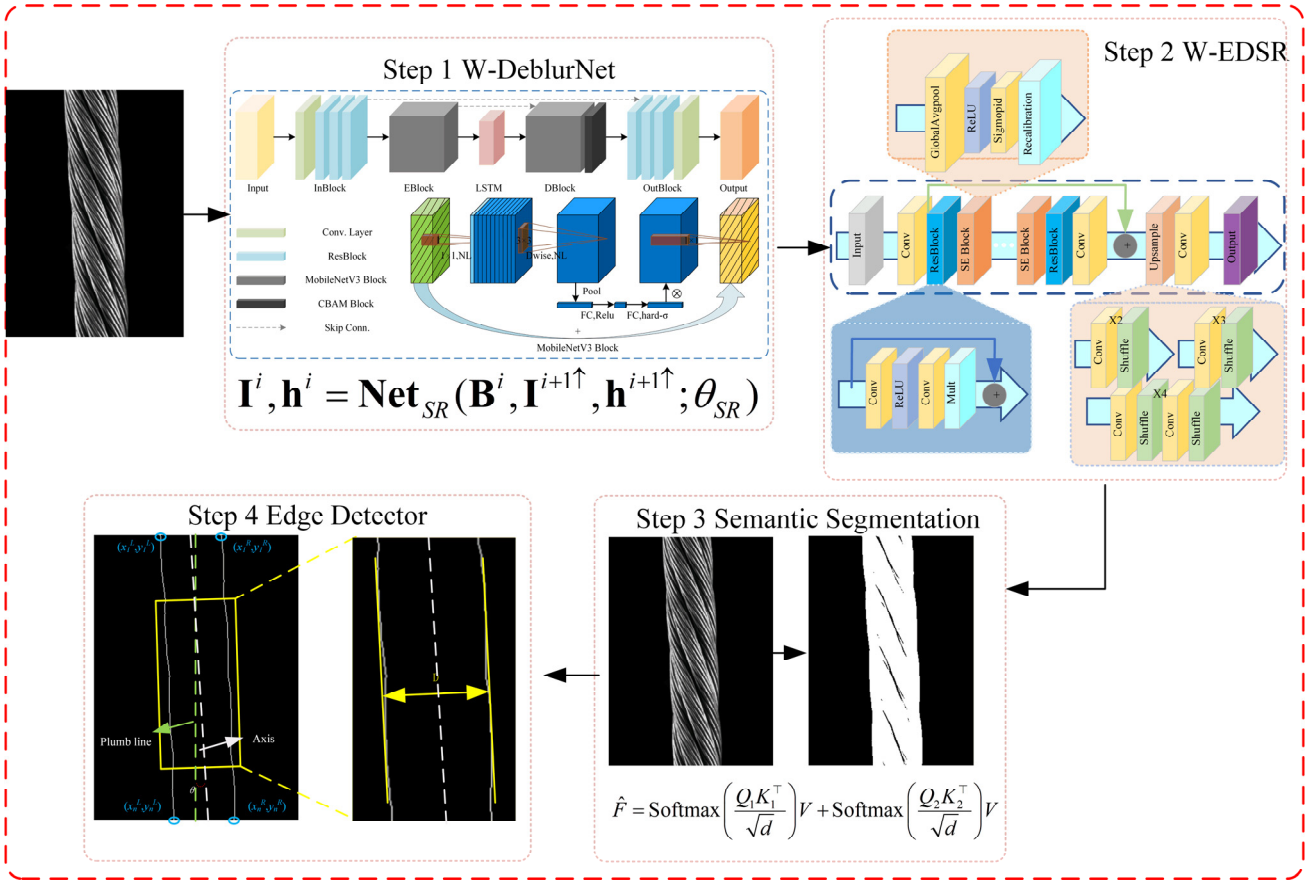

Figure 2. Architecture of the proposed method

The proposed method consists of four steps: image preprocessing, image enhancement, instance segmentation, and edge detection. The specific measurement process is as follows. First, a grayscale image with a length greater than

twice the strand pitch is captured using an industrial camera. This image is then fed into the improved W-DeblurNet network model for filtering, reducing the impact of speed variations on image quality during the steel wire rope's operation. Next, the processed image is input into a pre-trained W-EDSR network model to obtain a high Peak Signal-to-Noise Ratio (PSNR) steel wire rope image. The image is then subjected to the next step of semantic segmentation to facilitate the subsequent detection of the steel wire rope's edges. Finally, Canny edge detection is applied to identify the boundaries of the steel wire rope. The image is then straightened based on the rope's central axis using the least squares criterion, ensuring the steel wire rope appears vertical in the image. A section of the image with a length equivalent to twice the strand pitch is selected, and the distance between the left and right contours is calculated. This distance is then converted to the actual length, completing the steel wire rope diameter measurement. The research methodology is shown in Figure 2.

2.2. W-DeblurNet Network for Motion Blur Suppression

To address the problem of image degradation caused by motion and improper focusing during image acquisition, this study proposes a W-DeblurNet network for image deblurring. The network is an improved version of the original SRN-DeblurNet architecture [28], with enhanced deblurring capabilities. As a scale-recurrent network, its core idea is to progressively restore the latent sharp image through a multi-scale pyramid structure, optimizing structural details in blurred regions from coarse to fine. Compared with conventional multi-scale methods, W-DeblurNet shares the same sub-network structure across different scales, which significantly reduces the number of network parameters and improves both training and inference efficiency.

To enhance feature transmission across scales, convolutional long short-term memory (ConvLSTM) units are introduced. These units enable the hidden state to be recurrently passed between scales, allowing the network to adaptively model both the structural information of the image and the characteristics of the blur kernel. Specifically, at each scale, the network takes the blurry image at the current scale and the preliminary deblurring result upsampled from the previous scale as inputs to generate a clear latent image at the corresponding scale. The entire process can be formally represented as follows:

$$\mathbf{I}^i, \mathbf{h}^i = \mathbf{Net}_{SR}(\mathbf{B}^i, \mathbf{I}^{i+1\uparrow}, \mathbf{h}^{i+1\uparrow}; \theta_{SR}) \tag{1}$$

Where i denotes the scale index, and $i=1$ represents the finest scale. $\mathbf{B}^i$ and $\mathbf{I}^i$ refer to the blurry image and the estimated latent sharp image at scale i, respectively. $\mathbf{Net}_{SR}$ is the scale-recurrent network with θ_{SR} as its trainable parameters. Since the network adopts a recurrent structure, the hidden state varies across scales, carrying structural and kernel-related information from the previous coarser scale. $(*)^{\uparrow}$ denotes the adaptation operator that maps features or images from scale $i+1$ to scale i.

Considering the complex textures, fine structural details, and highly variable acquisition environments of mine steel wire rope images, the original SRN-DeblurNet network is insufficient to fully meet the requirements of real-time detection and practical deployment. Therefore, this paper introduces the following three improvements based on the original network to enhance its applicability and robustness:

1) The original encoder–decoder structure of the network is replaced with MobileNetV3 modules. MobileNetV3 offers a compact architecture and high computational efficiency, which significantly reduces the number of model parameters and computational complexity. This modification helps meet the real-time performance and deployment constraints commonly encountered in industrial field environments.

2) A channel attention module, CBAM (Convolutional Block Attention Module), is introduced during the decoding phase. CBAM adaptively enhances feature responses in both the channel and spatial dimensions, improving the model's ability to perceive subtle defects such as grease stains or burrs along the edges of the steel wire rope. This enhancement strengthens the fine-detail representation in the deblurring results.

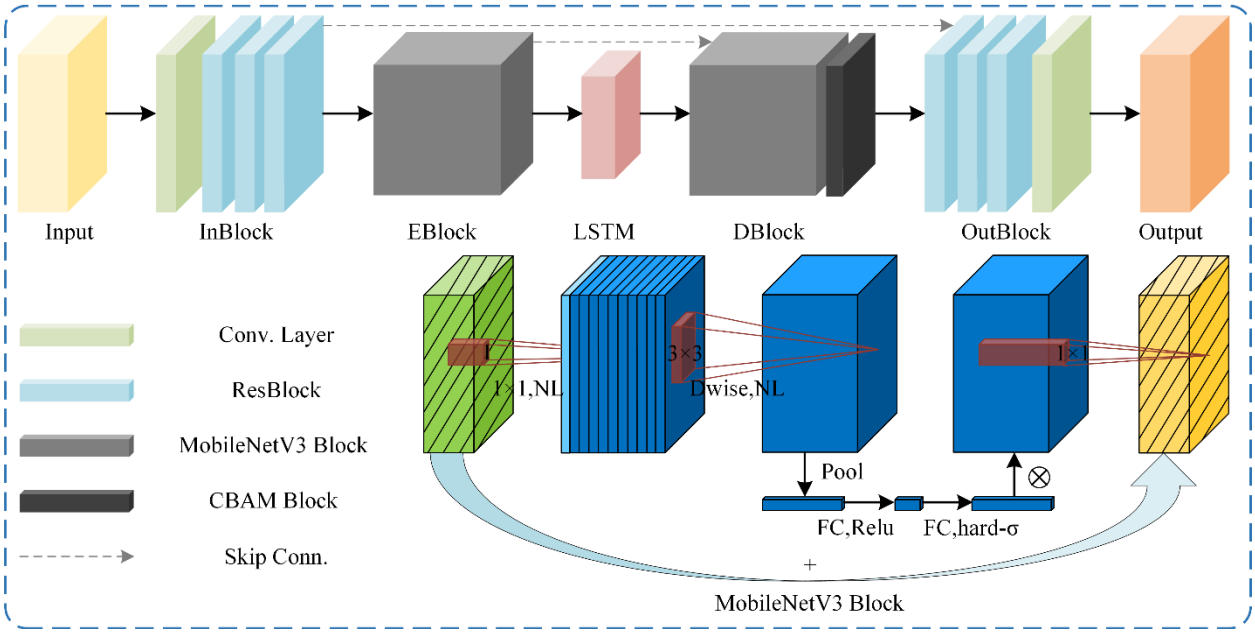

Figure 3. Network architecture of W-DeblurNet

3) A joint optimization strategy combining perceptual loss and edge loss is adopted during training. Perceptual loss leverages high-level semantic features extracted by a VGG network to improve the structural consistency and perceptual realism of the deblurred image. Edge loss emphasizes accurate reconstruction of image contours, which is particularly effective for enhancing the clarity of the steel wire rope's boundaries and local defect regions.

In summary, the proposed W-DeblurNet is designed to enhance deblurring capability while improving the preservation of critical structures and fine details. These improvements aim to make the network more suitable for real-time preprocessing of industrial steel wire rope images. By generating high-quality and reliable image inputs, the method is expected to benefit downstream tasks such as segmentation and defect detection. The overall network architecture is shown in Figure 3.

Furthermore, to suppress salt-and-pepper noise in the image and preserve the integrity of steel wire rope edges, we introduce an Adaptive Median Filter (AMF) at the output end of W-DeblurNet for fine-grained smoothing based on the deblurring results.

Median filtering is a typical nonlinear filtering technique with strong capabilities in removing salt-and-pepper noise while preserving edges, and it is widely used in image denoising tasks. The basic idea of median filtering is as follows: for a given pixel, a fixed-size window (e.g., 3×3 or 5×5) centered on the pixel is selected. The gray values of all pixels within the window are sorted, and the median value is taken as the new value for the central pixel. This operation can be expressed as:

$$I'(x,y) = \text{median}\{I(i,j) | (i,j) \in \text{N}(x,y)\} \quad (2)$$

Where $I(x,y)$ represents the pixel value at point (x,y) in the input image. $\text{N}(x,y)$ denotes the local neighborhood window centered at point (x,y). $I'(x,y)$ is the filtered pixel value at point (x,y) .Function $\text{median}\{*\}$ represents the median operation.

Compared to linear methods such as mean filtering, median filtering better preserves the edges and texture details of an image while smoothing noise. This method is used in this study to refine the boundary regions of the deblurring results produced by W-DeblurNet, enhancing image clarity and stability, and laying the foundation for subsequent instance segmentation and geometric feature extraction.

To validate the feasibility of this approach, a comparative test was conducted between W-DeblurNet and two commonly used deblurring networks on steel wire rope images: DeblurGAN and SRN-DeblurNet. After training all three networks on the same dataset, inference was performed on the same sample image. The evaluation metrics used were Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM). The inference results and evaluation scores are shown in Figure 4 and Table 1. A higher PSNR value indicates better image reconstruction quality, while an SSIM value closer to 1 suggests that the image structure is closer to the original image, reflecting better performance. As seen from the results, the proposed method not only shortens the running time but also ensures high image quality.

Table 1. Evaluation results of the W-DeblurNet network

Method	PSNR	SSIM	Time
DeblurGAN	30.274dB	0.876	0.07s
SRN-DeblurNet	32.617dB	0.870	0.13s
W-DeblurNet	32.163dB	0.896	0.06s

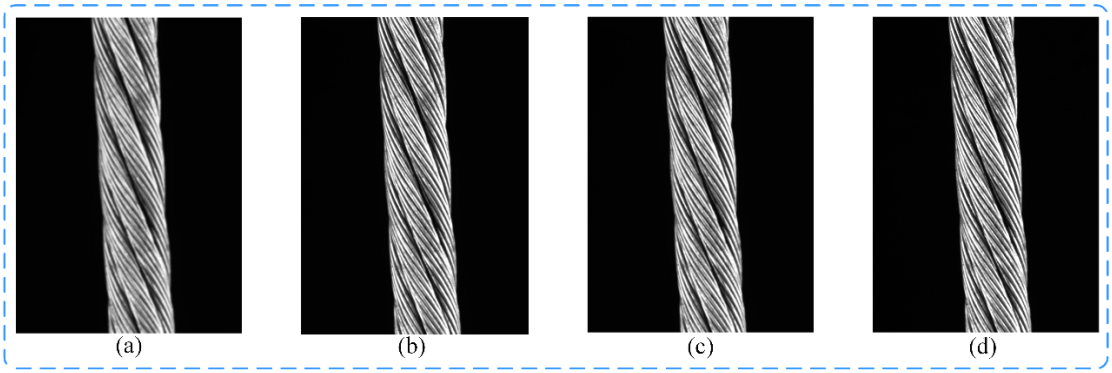

Figure 4. Comparison of deblurring networks. (a) Original image. (b) GAN-based methods DeblurGAN. (c) SRN-based deblurring network. (d) Improved W-DeblurNet

2.3. Image Enhancement based on W-EDSR Network

To enhance the model's ability to preserve structural information and improve detail perception in image reconstruction tasks, this paper proposes a W-EDSR network for image super-resolution. This method builds upon the classic EDSR (Enhanced Deep Super-Resolution Network) framework, incorporating a Squeeze-and-Excitation (SE) channel attention mechanism embedded in each residual block to achieve more effective feature enhancement. The network architecture is shown in Figure 5. Additionally, to further improve the image restoration quality, particularly the clarity of edge regions, a composite edge-aware loss function is designed to provide supervisory guidance during the training phase.

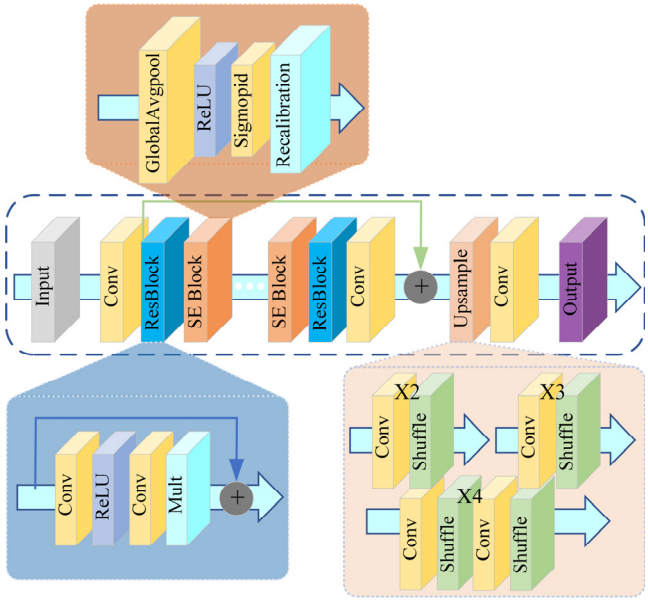

Figure 5. Network architecture of W-EDSR.

2.3.1. Model Architecture

The Squeeze-and-Excitation (SE) module is a lightweight channel attention mechanism designed to adaptively adjust the importance of each channel in the feature map, enhancing the representation of effective features. The specific process includes two stages: first, global average pooling is performed on each channel to obtain a channel descriptor vector (Squeeze), and then, a two-layer fully connected network generates the inter-channel weight coefficients (Excitation), which are normalized using the Sigmoid activation function. The formal representation is as follows:

$$s_c = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} x_c(i,j), \mathbf{w} = \sigma\left(W_2 \cdot \delta(W_1 \cdot \mathbf{s})\right) \quad (3)$$

Where is s_c the global descriptor for the c-th channel. $\mathbf{w}$ is the channel weight vector. $\delta(\cdot)$ denotes the ReLU activation function. $\sigma(\cdot)$ denotes the Sigmoid activation function. W_1 and W_2 are the weight matrices. The input feature map is weighted according to the channel weights to obtain the weighted output:

$$\tilde{x}_c = w_c \cdot x_c \quad (4)$$

Embedding the SE module into each residual block of EDSR helps suppress redundant features and enhances the model's response to key structures such as textures and edges. This improves feature utilization and reduces the interference from redundant channels, effectively lowering unnecessary computational overhead. Without significantly increasing the number of model parameters, it achieves dual improvements in image quality and runtime efficiency, significantly enhancing the detail quality of image restoration while ensuring computational efficiency.

2.3.2. Loss Function

In addition, to address the issues such as blurriness and structural incompleteness often exhibited by traditional L1 or L2 loss functions in restoring image edge regions, this paper designs a composite loss function that combines L1 loss with Gradient Difference Loss (GDL). This aims to enhance the model's perception and reconstruction capability of edge contours while improving overall pixel accuracy. The L1 loss is beneficial for enhancing the structural accuracy of the entire image and can effectively constrain the pixel-level error between the model's output and the ground truth image. It is defined as follows:

$$\mathrm{L}_{\mathrm{L1}} = \| I_{\mathrm{SR}} - I_{\mathrm{HR}} \|_1 \quad (5)$$

Where I_{SR} is the super-resolved image, and I_{HR} is the high-resolution ground truth image. GDL enhances edge information constraints by comparing the differences in image gradient variations. Its mathematical form is as follows:

$$\mathrm{L}_{\mathrm{GDL}} = \sum_{i,j} \left| \left| I_{\mathrm{HR}}^{i,j} - I_{\mathrm{HR}}^{i+1,j} \right| - \left| I_{\mathrm{SR}}^{i,j} - I_{\mathrm{SR}}^{i+1,j} \right| \right| + \left| \left| I_{\mathrm{HR}}^{i,j} - I_{\mathrm{HR}}^{i,j+1} \right| - \left| I_{\mathrm{SR}}^{i,j} - I_{\mathrm{SR}}^{i,j+1} \right| \right| \quad (6)$$

Therefore, the final composite loss function is defined as follows:

$$\mathrm{L}_{\mathrm{total}} = \mathrm{L}_{\mathrm{L1}} + \lambda \cdot \mathrm{L}_{\mathrm{GDL}} \quad (7)$$

Where λ is the balancing term that adjusts the contribution of the gradient loss in the overall training objective.

While effectively maintaining the overall structural consistency of the image, the proposed composite loss function significantly enhances the clarity and continuity of edge contours, thereby improving the perceptual quality and detail restoration of the image. This characteristic makes it particularly suitable for industrial inspection scenarios that demand high precision in identifying tiny defects and maintaining contour integrity. Compared with models that adopt a single L1 or L2 loss, the combined loss function demonstrates better convergence efficiency even at the early stage of training, reducing the number of iterations required for convergence, which facilitates faster model optimization and shortens the development cycle in industrial deployments.

To verify the effectiveness of the proposed method, the W-EDSR network was compared with four commonly used super-resolution models: Bicubic Interpolation, SRCNN, SRGAN, and EDSR. To ensure fairness and accuracy in testing, the same procedures were applied to all models: each was trained 1,000 times using the same dataset (including high-resolution and low-resolution image pairs). The inference results for the same input image are shown in Figure 6, and the evaluation metrics—PSNR and SSIM—are presented in Table 2. The results show that category-based learning methods (SRCNN, SRGAN, EDSR, and W-EDSR) outperform the interpolation-based Bicubic Interpolation method. Among the learning-based models, the W-EDSR network achieved the best PSNR and SSIM scores, with an inference time second only to the SRCNN network, making it the optimal choice for balancing speed and image quality.

2.4. Method for Measuring the Diameter of Steel Wire Ropes

In the field of industrial inspection, visual methods are often used to detect the diameter of objects, typically starting from the edges of the object to be measured. In order to obtain clear and continuous edges of the steel wire rope, this study first performs image segmentation processing on high-definition images of the steel wire rope. U-Net is a fully convolutional neural network architecture specifically

designed for image segmentation tasks. Due to its ability to effectively extract multi-level semantic features during the encoding process and reconstruct high-resolution spatial localization information during the decoding phase, U-Net has been widely applied in scenarios requiring precise segmentation, such as medical image analysis and industrial visual inspection [29].

Table 2. Evaluation results of the W-EDSR network

Method	PSNR	SSIM	Time
Bicubic	29.261dB	0.902	0.02s
SRCNN	29.974dB	0.916	0.05s
SRGAN	30.692dB	0.921	0.12s
EDSR	30.938dB	0.901	0.13s
W-EDSR	31.287dB	0.926	0.13s

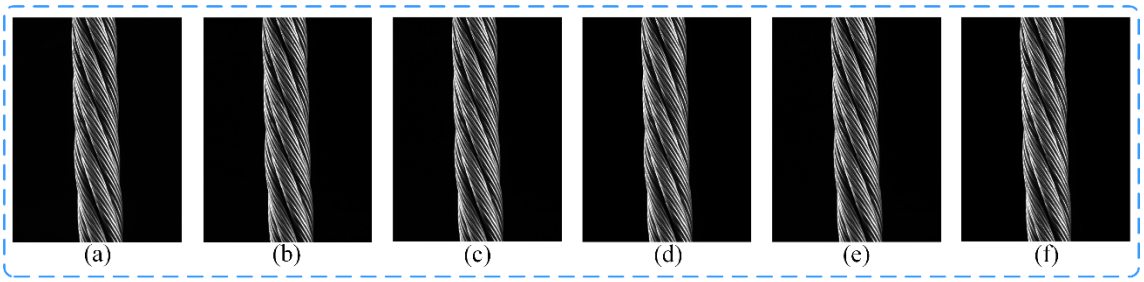

Figure 6. Comparison of super-resolution networks. (a) Original image. (b) Interpolation-based methods bicubic. (c) CNN-based methods SRCNN. (d) GAN-based methods SRGAN. (e) Attention mechanism-based approach EDSR. (f) Improved W-EDSR.

To address the challenge of segmenting the steel wire rope target against a complex background, this paper introduces an improved version of U-Net called RGBD-Unet. This network is mainly used for non-destructive testing of overhead steel wire ropes. It not only inherits the advantages of U-Net's symmetric encoder-decoder structure but also incorporates image depth information as an additional modality, effectively enhancing the model's ability to recognize targets in complex environments. RGBD-Unet adds an independent depth map encoding branch based on the original U-Net's single encoding path, emphasizing the collaborative learning capability between color and depth modalities, thereby improving the perception accuracy of the target contours. Furthermore, RGBD-Unet inserts a CMA (Cross-Modality Attention) module before each pooling layer, and then fuses information from different modalities in a weighted manner, thereby highlighting key regions and suppressing background interference. The fusion process can be summarized by the following formula:

$$\hat{F} = \underbrace{\text{Softmax}\left(\frac{Q_1 K_1^{\text{唯}}}{\sqrt{d}}\right)V_1}_{\text{Subspace attention output } F_1} + \underbrace{\text{Softmax}\left(\frac{Q_2 K_2}{\sqrt{d}}\right)V_2}_{\text{Subspace attention output } F_2} \quad (8)$$

Where $\hat{F}$ is the fusion result. Q_1 and Q_2 are the Query vectors extracted from the RGB and Depth feature. K_1 and K_2 are the Key vectors extracted from the RGB and Depth features. V_1 and V_2 represents the values in the input features, which are used to generate the weighted output features. $\text{Softmax}(\cdot)$ is the softmax function used to normalize the attention weights and highlight important feature positions. $Q_i K_i^{\bullet}$ is the result of the dot-product attention calculation, which computes the similarity between the Query vector and the Key vector. $\sqrt{d}$ is a scaling factor used to prevent the inner product results from becoming too large, causing instability.

By introducing a cross-modality attention mechanism in the feature extraction stage, RGBD-Unet achieves early and efficient fusion of color and depth information, enhancing the model's ability to perceive complex defects such as broken wires and twists in the steel wire rope. Experimental results show that this method outperforms traditional single-modality segmentation networks in terms of segmentation accuracy and robustness across multiple industrial datasets, especially maintaining good segmentation performance in scenarios with low contrast and strong noise, thus verifying the feasibility and effectiveness of the network structure in practical engineering applications.

To obtain clear and continuous contours of the steel wire rope, this paper uses the Canny edge detection algorithm to extract edge information from the steel wire rope images. The Canny operator is a multi-stage edge detection method that has good noise resistance and edge localization capability. The process mainly involves the following steps: First, the Canny operator smooths the image to reduce noise interference:

$$I_s(x,y) = G(x,y) * I(x,y) \quad (9)$$

Where $I_s(x,y)$ is the smoothed image. $G(x,y)$ represents the Gaussian filter. $I(x,y)$ is the original image. Subsequently, the operator calculates the gradient magnitude and direction at each pixel:

Where G_x and G_y are the gradients in the horizontal and vertical directions, respectively. After computing the gradients across the entire image, the Canny operator applies non-maximum suppression, retaining only the pixels with local maxima in the gradient direction as candidate edge points. Then, based on the set high and low thresholds, the pixels are classified as strong edges, weak edges, or non-edges. Strong edges are connected to adjacent weak edges,

forming a continuous edge structure.

$$M(x,y)=\sqrt{G_x^2(x,y)+G_y^2(x,y)},\quad \theta(x,y)=\arctan\left(\frac{G_y(x,y)}{G_x(x,y)}\right) \tag{10}$$

To accurately detect the diameter of the steel wire rope, it is essential to ensure that the rope is in a vertical position. However, due to lateral swinging during actual operation, the steel wire rope often appears tilted in the image. If the axis position is estimated using only two boundary points, significant errors are likely to occur. This paper proposes a highly robust method for rectifying the steel wire rope image, as shown in Figure 7(c). First, the coordinates of the left and right contour points of the detected steel wire rope in each image row are extracted and denoted as (x_i^L, y_i) and (x_i^R, y_i), where i=1,2,…,n,and n is the height of the image. Therefore, the midpoint coordinates of each row are given by:

$$(x_i^M, y_i^M)=\left(\frac{x_i^L+x_i^R}{2}, i\right) \tag{11}$$

A straight line is fitted to all the midpoints of the axis, and the fitted line equation is expressed as: $y=ax+b$. Where a represents the slope of the axis, and b is the intercept. To obtain the fitted line that best fits all the midpoint coordinates, the least squares criterion is applied to construct the sum of squared errors objective function:

$$E(a,b)=\sum_{i=1}^{n}\left(y_i-ax_i-b\right)^2 \tag{12}$$

Take the partial derivatives of this function with respect to a and b, and set the derivatives to zero. The optimal parameters are obtained as:

$$a=\frac{n\sum x_iy_i-\sum x_i\sum y_i}{n\sum x_i^2-(\sum x_i)^2}, b=\frac{\sum y_i-a\sum x_i}{n} \tag{13}$$

The final values of a and b can be used to construct the approximate straight line equation of the steel wire rope's axis. The slope a can also be used to calculate the tilt angle $\theta=\arctan a$ of the rope. Considering that the diameter of the steel wire rope needs to be measured as the distance between at least two adjacent outer tangents, a rectangular region with a width slightly larger than the nominal diameter and a length greater than twice the strand spacing is selected along the axis. This region is rotated to correct for the tilt angle θ. After the rope is vertical, the distance between the outer tangents of the left and right contours will give the diameter D, as shown in Fig. 7(d).

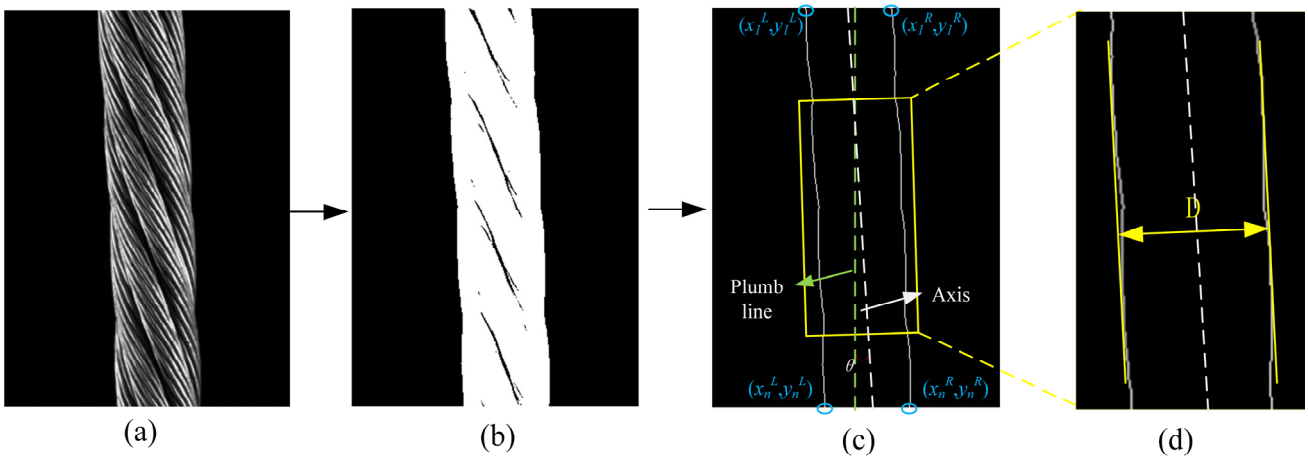

Figure 7. Diameter measurement method for steel wire ropes. (a) Enhanced image. (b) Image after segmentation. (c) Correct operation diagram. (d) Result diagram

3. Experimental Verification

3.1. Experimental Setup

To verify the effectiveness of the proposed method, both static and dynamic experiments were designed, with the following preparations made: Considering requirements such as image clarity, focal length, and lighting conditions, a 1.3-megapixel industrial camera was selected as the image acquisition device. It supports a frame rate of up to 210 fps and an exposure time of 20 μs, paired with a white ring-shaped industrial light source. The computer used for processing is configured as follows: Processor: 12th Gen Intel(R) Core (TM) i5-12490F; GPU: NVIDIA GeForce RTX 4070; RAM: 32 GB. The wire rope used in this study is a 6V×34-FC triangular-strand steel rope. Its structural characteristics are that each strand consists of 34 steel wires, for a total of 6 strands, with a fiber core (FC) as the rope core. The wire rope has a nominal diameter of 32 mm, a lay length of 220 mm, and a tensile strength of 1770 MPa.

3.2. Experimental Analysis

3.2.1. Static Experiment

The static experiment setup mainly consists of a shadowless light source, a computer system, an industrial camera, and the steel wire rope under test. During the equipment arrangement, a set of shadowless light sources is symmetrically placed on both sides of the steel wire rope to ensure consistent lighting conditions for image acquisition. The industrial camera is positioned directly facing the rope and mounted on a tripod, connected to the computer for image transmission, as shown in Figure 8.

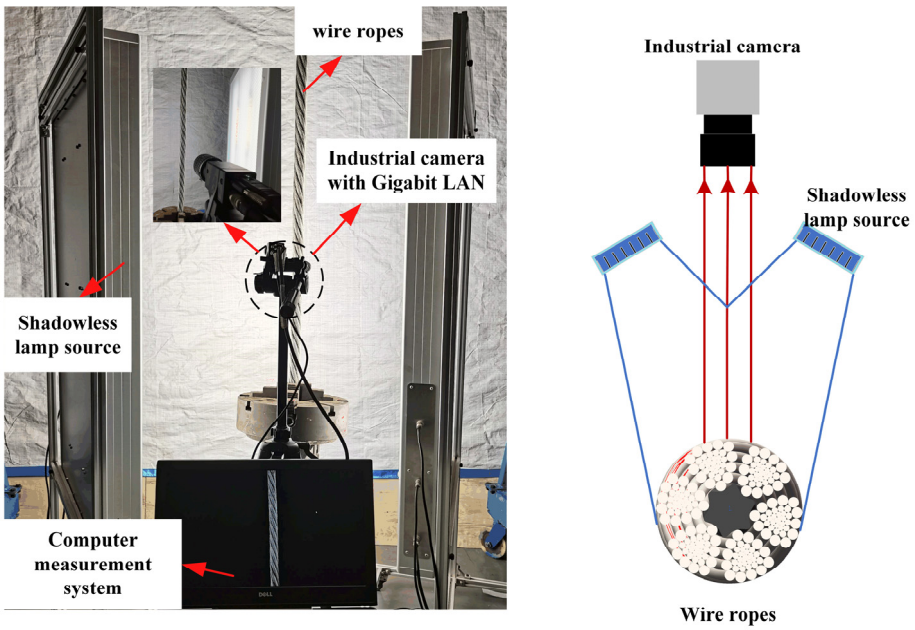

Figure 8. Static experiment equipment setup

After completing the setup of the experimental apparatus, a triangular calibration block was used to calibrate the steel wire rope. By attaching the calibration block to the rope, the actual width corresponding to each pixel in the image was calculated to be approximately 0.1436 mm. The calibration process is illustrated in Figure 9.

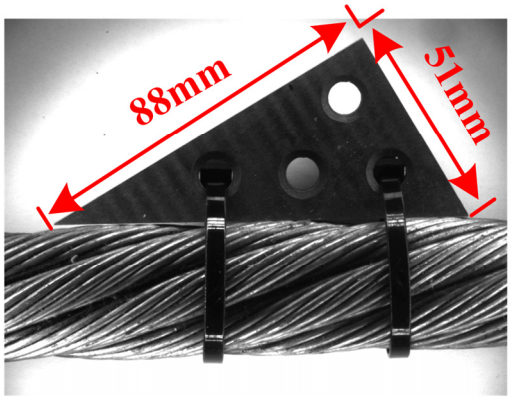

Figure 9. Calibration of pixel-to-distance ratio using a triangular calibration block

After calibration, images of the steel wire rope were captured. During image acquisition, photos were taken every 10 cm along the axial direction of the rope. Each group consisted of 10 images, and for each image, the rope diameter at the same position was also manually measured using a vernier caliper. The average of the manual measurements was recorded as the manual measurement value x_1, while the average of the values obtained through image processing was recorded as the algorithm measurement value x_2. After completing one group of data collection, the system was reset before proceeding to the next group. A total of 8 independent sample groups were collected.

During image processing, since the test object in this experiment was a static steel wire rope, deblurring was not strictly necessary. However, to maintain consistency with the dynamic experiment procedure, deblurring was still applied using a neural network. The raw images were then processed with median filtering before being enhanced using the improved W-EDSR network. Image segmentation was performed using the RGBD-Unet, and the edges of the rope were extracted using the Canny operator. Finally, image rectification and diameter measurement were completed. The results are shown in the table 3. Here, the manual measurement value x_1 refers to the average value obtained from manual measurements with a vernier caliper. The algorithm measurement value x_2 is the average value obtained using the proposed method.

Table 3. Results of the static experiment for steel wire rope diameter measurement

Experimental Group number	Manual measurements x_1/mm	Algorithm measurements x_2/mm	Dimensional deviation value Δ_x/mm	Maximum deviation value Δ_{xm}/mm	Standard deviation σ/mm	Accuracy η/%
1	29.62	29.832	0.212	0.269	0.0509	99.29
2	29.58	29.741	0.161	0.227	0.0626	99.46
3	29.64	29.845	0.205	0.291	0.0318	99.31
4	29.64	29.479	0.161	0.251	0.0416	99.46
5	29.46	29.614	0.114	0.306	0.0753	99.48
6	29.56	29.788	0.228	0.275	0.0437	99.23
7	29.58	29.767	0.187	0.374	0.0513	99.37
8	29.72	29.471	0.129	0.238	0.0564	99.16
Average value	29.60	29.692	0.173	0.279	0.0517	99.35

The dimension deviation Δx is the absolute difference between the two averages and is used to calculate the accuracy. The maximum deviation Δx_m, also reported as an absolute value, evaluates the robustness of the method. The standard deviation σ and accuracy η are used to assess the dispersion of the data and the precision level of the algorithm, respectively. Accuracy is defined as:

$$\eta = \left(1 - \frac{\Delta x}{x_1}\right) \times 100\% \tag{14}$$

3.2.2. Dynamic Experiment

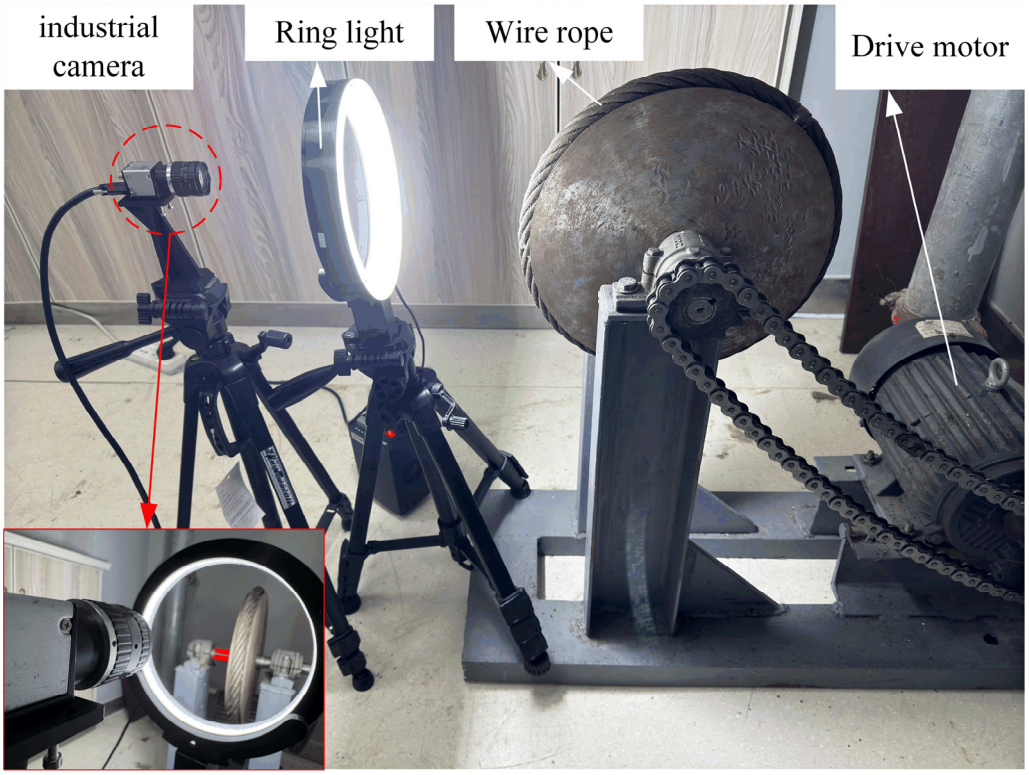

Figure 10. Dynamic experiment equipment setup

The dynamic experimental setup is shown in Figure 10, and it mainly consists of an image acquisition module, an electrical control module, and the segment of steel wire rope under test. The steel wire rope is wound around a rotating disk, with both ends welded together to form a closed-loop structure. When the three-phase asynchronous motor operates, it drives the rotating disk via a chain transmission system, which in turn causes the steel wire rope to rotate, thereby simulating the lifting motion of the rope. The image acquisition module comprises an industrial camera for image capture, a ring-shaped light source for illumination, and a tripod for mounting and stability. The electrical control module includes the three-phase asynchronous motor used to drive the rope and an electrical cabinet responsible for power supply and frequency control. The industrial camera, ring light source, and steel wire rope are aligned in a straight line, with the optical axis of the camera adjusted to be perpendicular to the rope segment and the target positioned at the center of the field of view.

After re-calibration, the dynamic experiment was conducted. For each experimental group, ten points were manually measured along one full revolution of the stationary steel wire rope. The average of these measurements was recorded as the manual measurement value x_1. After the rope began to move, the algorithm was used to measure the diameter of the rope segment in the image in real time. Each captured image covered approximately 15 cm of rope length. When performing batch inference every 5 frames, with partial module optimization and batch processing on an NVIDIA GeForce RTX 4070, the total inference time was measured to be 0.1379 seconds. This indicates that the model is capable of detecting the rope diameter at a motion speed of 10 m/s. Based on the fixed outer diameter $d = 320mm$ of the steel wire rope, the disk's rotational speed was adjusted to 10 rad/s to match the test conditions, and a total of 8 independent groups were collected for the experiment.

The experimental results are shown in Table 4.

Based on the results of both experiments, it can be observed that when using the same model of steel wire rope for measurement, the average size deviation of the algorithm's measurements relative to the true value in the stationary state is 0.173 mm. The average maximum deviation is 0.279 mm, and the average accuracy is 99.35%, demonstrating the high detection accuracy of the method.

When the steel wire rope is moving at a speed of 10 m/s, the average size deviation increases to 0.227 mm, the average maximum deviation is 0.325 mm, and the average accuracy is 99.24%. Although the accuracy is slightly lower compared to the static experiment results, it still meets the design requirements of the detection method.

3.3. Ablation Experiment

To verify the effectiveness of the proposed method, this study designed four sets of ablation experiments, including: the baseline model EDSR, W-EDSR, W-DeblurNet, and W-EDSR + W-DeblurNet, which combines the two

improvements. As shown in Table 5.

Table 4. Results of the dynamic experiment for steel wire rope diameter measurement

Experimental Group number	Manual measurements x_1/mm	Algorithm measurements x_2/mm	Dimensional deviation value Δx/mm	Maximum deviation value Δx_m/mm	Standard deviation σ/mm	Accuracy η/%
1	29.46	29.683	0.223	0.362	0.0477	99.25
2	29.64	29.421	0.219	0.297	0.0724	99.26
3	29.62	29.812	0.192	0.336	0.0492	99.36
4	29.68	29.867	0.187	0.374	0.0441	99.37
5	29.54	29.841	0.301	0.263	0.0589	98.99
6	29.66	29.439	0.221	0.408	0.0617	99.25
7	29.62	29.878	0.258	0.241	0.0323	99.14
8	29.72	29.932	0.212	0.315	0.0629	99.29
Average value	29.62	29.734	0.227	0.325	0.0537	99.24

First, compared with the baseline EDSR, W-EDSR achieved a significant improvement in reconstruction quality, with PSNR increasing from 30.938 dB to 32.287 dB and SSIM increasing from 0.901 to 0.926. Although no acceleration in inference time was achieved (both were 0.13 s), the performance improvement brought by the enhancement module is more significant. Second, using only W-DeblurNet, the PSNR reached 32.163 dB and the SSIM was 0.896. While the overall performance was better than the baseline EDSR, it was slightly lower than W-EDSR. Finally, combining W-EDSR with W-DeblurNet further improved model performance, achieving a PSNR of 32.461 dB, the highest among the four experimental groups. Although the inference time increased from 0.13 s and 0.06 s for single modules to 0.15 s, this increase is reasonable and acceptable. Meanwhile, the SSIM was 0.925, essentially consistent with W-EDSR; although slightly lower, the difference was negligible and its impact on overall structural information could be considered insignificant. In summary, the two improved modules are complementary, and their combined use effectively improves reconstruction performance.

Table 5. Evaluation results of the W-EDSR network

Method	PSNR	SSIM	Time
EDSR	30.938dB	0.901	0.13s
W-EDSR	32.287dB	0.926	0.13s
W-DeblurNet	32.163dB	0.896	0.06s
W-EDSR+ W-DeblurNet	32.461dB	0.925	0.15s

4. Conclusion

This paper proposes a steel wire rope diameter detection method based on super-resolution (SR) to meet the requirements of online monitoring in mine hoisting systems. The method utilizes an RGBD camera to capture images of the steel wire rope, which are then transmitted to a computer. By combining W-DeblurNet and W-EDSR, the boundary clarity of the original grayscale images is significantly enhanced. Smooth and continuous rope edges are successfully extracted using a semantic segmentation algorithm fused with depth information, along with an edge detection algorithm. Additionally, a steel rope rectification method based on the least squares criterion is proposed to address the challenge of measuring inclined ropes. Experimental results show that the proposed method achieves an accuracy of 99.35% for static rope detection and 99.24% for ropes moving at a speed of 10 m/s, both exceeding the performance of traditional measurement approaches. Moreover, the method demonstrates strong robustness against motion blur. Therefore, this method enables fast, accurate, and non-contact online detection of steel wire ropes in mine hoisting operations, significantly improving the efficiency and quality of rope inspection.

Acknowledgments

This work was supported in part by the Major Science and Technology Project of Anhui Provincial Science and Technology Innovation Platform (No.S202305a12020036), the Engineering Laboratory for Safe and Precise Coal Mining of Anhui Province (No.ESCMP202406), Anhui Key Laboratory of Mine Intelligent Equipment and Technology, Anhui University of Science & Technology (No. ZKSYS 202405), the Scientific Research Foundation for High Level Talents of Anhui University of Science and Technology (No.2025yjrc0073).

References

[1] T. Wang and V. J. L. Gan, "Automated joint 3D reconstruction and visual inspection for buildings using computer vision and transfer learning," *Automation in Construction*, vol. 149, p. 104810, May 2023, doi: 10.1016/j.autcon.2023.104810.

[2] Y. D. V. Yasuda, F. A. M. Cappabianco, L. E. G. Martins, and J. A. B. Gripp, "Aircraft visual inspection: A systematic literature review," *Computers in Industry*, vol. 141, p. 103695, Oct. 2022, doi: 10.1016/j.compind.2022.103695.

[3] L. Ren, Z. Liu, and J. Zhou, "Shaking Noise Elimination for Detecting Local Flaw in Steel Wire Ropes Based on Magnetic Flux Leakage Detection," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–9, 2021, doi: 10.1109/TIM.2021.3112792.

[4] H. Wang, Q. Li, S. Han, P. Li, J. Tian, and S. Zhang, "Wire Rope Damage Detection Signal Processing Using K-Singular Value Decomposition and Optimized Double-Tree Complex Wavelet Transform," *IEEE Transactions on Instrumentation*

and Measurement, vol. 71, pp. 1–12, 2022, doi: 10. 1109/ TIM. 2022. 3216670.

[5] J. Tian, C. Zhao, W. Wang, and G. Sun, "Detection Technology of Mine Wire Rope Based on Radial Magnetic Vector With Flexible Printed Circuit," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–10, 2021, doi: 10.1109/TIM.2021.3096288.

[6] Y. L. Guo, G. X. Wu, X. L. Liu, and X. L. Xu, "REVIEW OF FAULT DIAGNOSIS METHODS FOR ROTATING MACHINERY BASED ON DEEP LEARNING," in *The 8th International Symposium on Test Automation & Instrumentation (ISTAI 2020)*, Nov. 2020, pp. 175–180. doi: 10.1049/icp.2021.1316.

[7] K. Qiu, L. Tian, and P. Wang, "An Effective Framework of Automated Visual Surface Defect Detection for Metal Parts," *IEEE Sensors Journal*, vol. 21, no. 18, pp. 20412–20420, Sep. 2021, doi: 10.1109/JSEN.2021.3095410.

[8] X. Li, J. Li, Y. Qu, and D. He, "Semi-supervised gear fault diagnosis using raw vibration signal based on deep learning," *Chinese Journal of Aeronautics*, vol. 33, no. 2, pp. 418–426, Feb. 2020, doi: 10.1016/j.cja.2019.04.018.

[9] Y. Zhang, G. Cao, and J. Cao, "Target-less approach of wire rope rotation measurement," *Measurement*, vol. 221, p. 113489, Nov. 2023, doi: 10.1016/j.measurement.2023.113489.

[10] A. Kazerouni, S. Karimijafarbigloo, R. Azad, Y. Velichko, U. Bagci, and D. Merhof, "Fusenet: Self-Supervised Dual-Path Network For Medical Image Segmentation," in *2024 IEEE International Symposium on Biomedical Imaging (ISBI)*, May 2024, pp. 1–5. doi: 10.1109/ISBI56570.2024.10635112.

[11] A. Albanese, M. Nardello, G. Fiacco, and D. Brunelli, "Tiny Machine Learning for High Accuracy Product Quality Inspection," *IEEE Sensors Journal*, vol. 23, no. 2, pp. 1575–1583, Jan. 2023, doi: 10.1109/JSEN.2022.3225227.

[12] P. Zhou, G. Zhou, S. Wang, H. Wang, Z. He, and X. Yan, "Visual Sensing Inspection for the Surface Damage of Steel Wire Ropes With Object Detection Method," *IEEE Sensors Journal*, vol. 22, no. 23, pp. 22985–22993, Dec. 2022, doi: 10.1109/JSEN.2022.3214109.

[13] X. Wang and Z. Kan, "Defect Detection of Steel Wire Rope in Coal Mine Based on Improved YOLOv5 Deep Learning," *Journal of Information Processing Systems*, vol. 19, no. 6, pp. 745–755, doi: 10.3745/JIPS.04.0293.

[14] Y. Dong, Y. Pan, D. Wang, and T. Cheng, "Corrosion detection and evaluation for steel wires based on a multi-vision scanning system," *Construction and Building Materials*, vol. 322, p. 125877, Mar. 2022, doi: 10.1016/j.conbuildmat. 2021.125877.

[15] P. Zhou, G. Zhou, H. Wang, D. Wang, and Z. He, "Automatic Detection of Industrial Wire Rope Surface Damage Using Deep Learning-Based Visual Perception Technology," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–11, 2021, doi: 10.1109/TIM.2020.3011762.

[16] J. Yu, X. Cheng, and Q. Li, "Surface Defect Detection of Steel Strips Based on Anchor-Free Network With Channel Attention and Bidirectional Feature Fusion," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–10, 2022, doi: 10.1109/TIM.2021.3136183.

[17] A. Assadzadeh, M. Arashpour, I. Brilakis, T. Ngo, and E. Konstantinou, "Vision-based excavator pose estimation using synthetically generated datasets with domain randomization," *Automation in Construction*, vol. 134, p. 104089, Feb. 2022, doi: 10.1016/j.autcon.2021.104089.

[18] Y. Wang, J. Luo, C. Liu, X. Yuan, K. Wang, and C. Yang, "Layer-Wise Residual-Guided Feature Learning With Deep Learning Networks for Industrial Quality Prediction," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–11, 2022, doi: 10.1109/TIM.2022.3214611.

[19] L. Wu *et al.*, "Deep learning-based super-resolution with feature coordinators preservation for vision-based measurement," *Structural Control and Health Monitoring*, vol. 29, no. 12, p. e3107, 2022, doi: 10.1002/stc.3107.

[20] X. Ge, H. Cui, Z. Xu, M. He, and X. Han, "Super-Resolution Image Reconstruction Method for Micro Defects of Metal Engine Blades," *gxxb*, vol. 43, no. 2, p. 0210001, Feb. 2023, doi: 10.3788/AOS221263.

[21] L. Schermelleh *et al.*, "Super-resolution microscopy demystified," *Nat Cell Biol*, vol. 21, no. 1, pp. 72–84, Jan. 2019, doi: 10.1038/s41556-018-0251-8.

[22] Y. Wang *et al.*, "Remote sensing image super-resolution and object detection: Benchmark and state of the art," *Expert Systems with Applications*, vol. 197, p. 116793, Jul. 2022, doi: 10.1016/j.eswa.2022.116793.

[23] C.-Q. Feng, B.-L. Li, Y.-F. Liu, F. Zhang, Y. Yue, and J.-S. Fan, "Crack assessment using multi-sensor fusion simultaneous localization and mapping (SLAM) and image super-resolution for bridge inspection," *Automation in Construction*, vol. 155, p. 105047, Nov. 2023, doi: 10.1016/j. autcon. 2023.105047.

[24] L. Chen, K. Meng, H. Zhang, J. Zhou, and P. Lou, "SR-FABNet: Super-Resolution branch guided Fourier attention detection network for efficient optical inspection of nanoscale wafer defects," *Advanced Engineering Informatics*, vol. 65, p. 103200, May 2025, doi: 10.1016/j.aei.2025.103200.

[25] G. Wang *et al.*, "Efficient multi-branch dynamic fusion network for super-resolution of industrial component image," *Displays*, vol. 82, p. 102633, Apr. 2024, doi: 10.1016/j. displa. 2023. 102633.

[26] X. Sun *et al.*, "A Multiscale Attention Mechanism Super-Resolution Confocal Microscopy for Wafer Defect Detection," *IEEE Transactions on Automation Science and Engineering*, vol. 22, pp. 1016–1027, 2025, doi: 10.1109/ TASE. 2024. 3358 693.

[27] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced Deep Residual Networks for Single Image Super-Resolution," in *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Jul. 2017, pp. 1132–1140. doi: 10.1109/CVPRW.2017.151.

[28] X. Tao, H. Gao, X. Shen, J. Wang, and J. Jia, "Scale-Recurrent Network for Deep Image Deblurring," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Jun. 2018, pp. 8174–8182. doi: 10.1109/ CVPR. 2018.00853.

[29] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds., Cham: Springer International Publishing, 2015, pp. 234–241. doi: 10.1007/978-3-319-24574-4_28.