IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 2 January 2024, accepted 31 January 2024, date of publication 8 February 2024, date of current version 22 February 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3363834

RESEARCH ARTICLE

Wear State Detection of Conveyor Belt in Underground Mine Based on Retinex-YOLOv8-EfficientNet-NAM

LIJIE YANG, GUANGYU CHEN, JIEHUI LIU, AND JINXI GUO
School of Mechanical and Equipment Engineering, Hebei University of Engineering, Handan 056000, China

Corresponding author: Jiehui Liu (jiehuiliu@hebeu.edu.cn)

This work was supported in part by the Laboratory of Intelligent Industrial Equipment Technology of Hebei Province, Hebei University of Engineering; in part by the Hebei Natural Science Foundation under Grant E2019402436; and in part by the Handan Science and Technology Bureau Project under Grant 21422907300.

ABSTRACT The belt surface of the mine belt conveyor can cause serious wear under the condition of long-term high-load operation, which can have a negative impact on production, bring economic losses, even endanger personal safety, and cause serious production accidents. Manual detection requires a lot of manpower and material resources, and is highly dependent on empirical judgment, which is with low efficiency and security risks. Therefore, in this study, we introduce a new conveyor belt wear detection algorithm Retinex-YOLOv8-EfficientNet-NAM (RYEN algorithm) based on deep learning and machine vision technology to replace manual detection, improving detection efficiency and recognition accuracy. The wear degree of belt is reclassified and defined according to the mechanical properties and wear texture characteristics of belt with different wear degrees, and a new special data set for belt wear detection is established. Aiming at the low brightness, high noise and complex working conditions of the underground mine, Gaussian filtering and bilateral filtering are used as the central surround function of the improved Retinex algorithm, and then channel fusion is performed with the image after histogram equalization and adaptive brightness adjustment. The improved Retinex multi-image fusion algorithm is used to preprocess the collected image. EfficientNet has the performance of reasonably allocating the input resolution, network depth, and channel width, and can maximize the performance of the network with limited resources. EfficientNet is used to replace Darknet53 of YOLOv8 as the backbone of the feature extraction network, which improves the detection accuracy under limited computing resources. A lightweight attention module NAM is added to the improved network, which improves the detection speed without reducing the detection accuracy. Experimental results show that RYEN algorithm effectively maintains the smoothness of the image during the image preprocessing stage, improves the brightness and contrast of the image, and better preserves the edge information of the image. RYEN algorithm achieves a detection speed of 66FPS and an average accuracy of 98.57%. Compared with the original YOLOv8 algorithm, the accuracy of RYEN algorithm is increased by 6.4% and the speed is increased by 13.2%. In comparison experiments with similar methods, RYEN algorithm occupies less hardware resources, has strong generalization ability, good performance, and has high detection speed and accuracy.

INDEX TERMS Conveyor belt wear inspection, Retinex-YOLOv8-EfficientNet-NAM, machine vision, deep learning, YOLOv8.

I. INTRODUCTION

As one of the main conveying equipment under the mine, the stable operation of the mining belt conveyor is the key to efficient production. At present, mining conveyor is developing in the direction of long-distance, high efficiency, large capacity, and intelligentization [1].

The conveyor belt as an important part of the belt conveyor, mainly serves the role of carrying materials and transferring transportation [2]. The conveyor belt is vulnerable to foreign

The associate editor coordinating the review of this manuscript and approving it for publication was Byung Cheol Song.

bodies which can be extremely destructive. There are mainly two types of situations that can cause damage to conveyor belts. One is that when the roller is broken or the material is blocked, the conveyor belt often suffers from longitudinal tears, surface scratches, and other damages. The other is fatigue wear caused by long-term operation. As the conveyor belt wear intensifies, the cotton canvas [4] and steel wire rope inside is gradually exposed to the skin [5], and chemical reactions occur when they come into contact with air and coal slurry. If it cannot be detected and treated effectively in time, the internal material of the conveyor belt can cause corrosion, and the damage may further expand, resulting in belt breakage and other accident.

The commonly used conveyor belt damage detection methods can be divided into manual detection methods and non-destructive detection methods. The main methods used for manual inspection include observation of surface damage on conveyor belts, belt speed detection, tension detection, etc., which require a large amount of work [6]. Therefore, manual inspection requires a great effort, and the recognition accuracy is unsatisfactory. These methods need to be operated in the state of static offline, which makes it difficult to meet the actual detection requirements. There are various non-destructive testing (NDT) methods, such as tear sensors, eddy current detection [7], weak magnetic detection [8], X-ray detection [9], ECD (Embedded Conductive Detection) technology [10], etc. Most of them can meet the needs of real-time detection, but these methods still have problems such as low sensitivity, radiation to the human body, and high price.

In recent years, the rapid development of computer technology has provided strong support for the development of machine vision technology. The maturity of machine vision technology has been gradually applied in various fields and has been successfully applied in the field of belt conveyor. For example, the measurement of coal volume and coal flow on the conveyor belt [11], classification of materials [12], on-line analysis of ore composition [13], [14], [15], [16], [17], gangue identification [18], [19], [20], conveyor belt detection [21], [22], [23], and so on. Zhang et al. [24] proposed a conveyor belt deviation detection method based on machine vision technology for fast and timely detection of conveyor belt runout status. Jose et al. [25] used a combination of machine vision techniques and multiple real-time sensor data to inspect conveyor belt system health. Hao et al [26] proposed a multi-class support vector machine detection system based on visual saliency. It is used to effectively extract grey values and identify and classify damage locations under dry and wet working conditions. Lv et al [27] proposed an improved grey scale centre of gravity method (IGGM). The features of longitudinal belt tear are transformed into linear features of line laser and belt tear detection is performed based on the features of laser centreline. Qiao et al. [28] used a combination of visible light and infrared vision to detect the longitudinal tearing of conveyor belts in real time. This fusion method greatly improved the accuracy of tear detection. Hou et al. [29] used multispectral vision detection to identify longitudinal belt tears while also classifying other states of the belt. Guo et al. [30], [31] used binarisation to process the images and combined with the collected sound signals to make a comprehensive judgment on the belt status. The theory of belt damage detection is summarised on the basis of machine vision and deep learning, which provides theoretical support for related research. Guo et al. [32] proposed a multi-classification condition method based on the generative adversarial network using deep learning to effectively detect belt surface damage. Liu et al. [33] proposed a belt damage detection method by fusing image spatiotemporal features, which effectively improved the accuracy of detection. Wang et al. [34] proposed a foreign object recognition method for conveyor belts based on improved SSD algorithm to fundamentally prevent the occurrence of tearing.

In summary, the current research on conveyor belt inspection mainly focuses on belt damage detection, with less research on belt wear. Until now, the main detection method still relies on manual inspection, and the existing manual inspection requires a lot of manpower and material resources in the work. The inspection process is highly dependent on the inspector's empirical judgement, which results in low inspection efficiency and potential safety hazards. A new belt wear detection method of Retinex-YOLOv8-EfficientNet-NAM (RYEN algorithm) in this paper is proposed for conveyor belt wear state detection in underground mines based on deep learning and machine version. RYEN algorithm is to replace manual inspection to achieve a more reliable, more effective, and more intelligent belt wear detection. The main contributions of this paper are summarized as follows.

- Compared with most current conveyor belt damage detection methods, the RYEN algorithm based on deep learning can achieve real-time detection of conveyor belt wear status, greatly improving detection efficiency and accuracy.
- The classification of conveyor belt wear degree is redefined and classified based on mechanical property data support and wear texture characteristics, which makes the classification more detailed, and maximizes the utilization of conveyor belt.
- The RYEN algorithm in this paper integrally considers the effect of model scaling on the detection results and achieves the best speed-accuracy trade-off on the dataset, which can also be applied to other target detection in other fields.

II. CLASSIFICATION OF BELT WEAR BASED ON MECHANICAL TEST PROPERTIES AND WEAR TEXTURE CHARACTERISTICS

As the main loss component of conveyor belts in daily operation, the performance of conveyor belts will also change with the deepening of their wear. The full layer tensile strength of the conveyor belt is a key performance parameter in the quality inspection of the conveyor belt [34]. If the strength does not meet the requirements, the conveyor belt

FIGURE 1. Conveyor belt wear types.

may inevitably tear or even break when the dynamic tension changes significantly.

The tensile test is carried out on the conveyor belt using a tensile testing machine. NN-100 nylon belt commonly used in conveyor belts is selected as the test sample, as shown in Fig. 1. According to the national standard GB/T 3690-2017, the full thickness tensile test specimens are prepared. Conveyor belt cutting samples with different wear degrees are selected. The samples are exposed in the environment with the temperature of 23 ± 2°C and the humidity of 50% ± 5% for three days to allow the residual stress generated by the cutting of the conveyor belt to self-regulate and release. The sample width is 50mm, the effective tensile length is 300mm, the experimental cutting length is 500mm. The clamping length in this experiment is 100mm, not less than 50mm stipulated by the national standard. This helps to increase the contact area between the fixture and the conveyor belt, increase the clamping force, and reduce the dislocation or pulling of the clamping force and the belt between the layers during the stretching process. The selected tensile testing machine for the experiment is WAW-100, as shown in Fig.2. The jaw type is Q-type, and the stretching speed of the tester gripper is 100mm/min, with a clamping length of 100mm.

The specific structure of the conveyor belt, as shown in Fig.3, mainly consists of a nylon core skeleton and a laminated covering layer. The performance results of the conveyor belt test after the tensile test are shown in Fig.4. The joints between the layers of the conveyor belt have no wear and are relatively tight after one layer of wear, and the tensile performance curve is relatively smooth. From the structural characteristics of the conveyor belt, due to the undamaged rubber layer on the surface, the surface of the undamaged conveyor belt is smooth. The rubber layer on the surface of the wear conveyor belt breaks down, exposing the nylon core layer. The nylon core exhibits a distinctive wear boundary that borders the rubber layer, providing a unique wear textured appearance. The elongation distance at the final

FIGURE 2. WAW-100 type tensile testing machine.

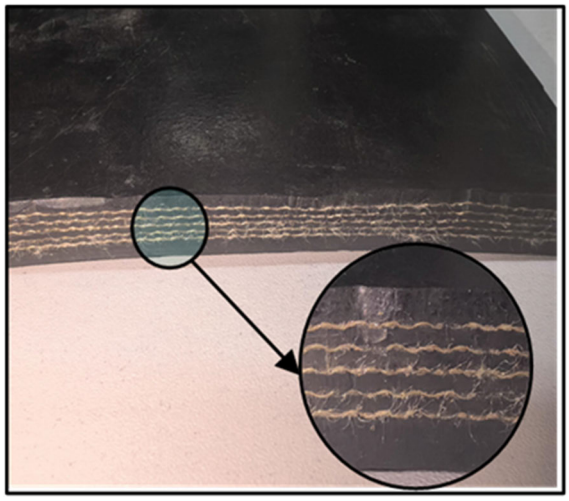

FIGURE 3. Conveyor belt internal structure.

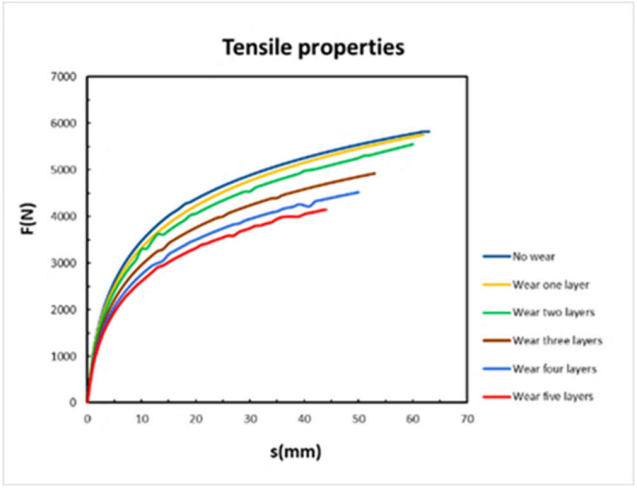

FIGURE 4. Tensile test results.

fracture is between 60 mm and 65 mm. The characteristic of the wear texture from the second to fifth layers is that

as wear increases, the rubber and nylon layers are exposed deeper, thereby showing more wear boundaries. As the degree of wear increases, the resin layer and nylon core layer of conveyor belt have a certain relaxation, which complicates the internal mechanical properties. Tests have shown that for two-layer wear, three-layer wear, four-layer wear, and five-layer wear, the elongation distance at the final fracture point is roughly distributed between 55mm-60mm, 50mm-55mm, 45mm-50mm, and 40mm-45mm. At the same time, the maximum tensile force that the conveyor belt can withstand is also decreasing.

According to standard GB/T 3690-2017 implemented by the conveyor belt, the longitudinal tensile elongation of the full thickness is not less than 10%. Therefore, when the wear level of the conveyor belt reaches three layers, it exceeds the allowable range of standard conveyor belt performance and poses a safety hazard. It should be replaced promptly.

The tensile test shows that the number of wear layers directly affects the tensile properties of the conveyor belt, and the different wear layers have unique textures. Thus, the unique wear texture of conveyor belts with different wear layers combined with the tensile strength of each wear level obtained from tensile performance tests are used as the categorization criterion. A category threshold is set according to the criterion, and when the tensile performance of the worn conveyor belt exceeds the threshold, the conveyor belt is regarded as end-of-life and replaced in time.

Therefore, the wear detection is classified according to the number of layers of the conveyor belt, which is divided into no wear, one layer wear, two layers wear, three layers wear, four layers wear and five layers wear.

III. WEAR DETECTION METHOD OF RETINEX-YOLOv8-EFFICIENTNET-NAM (RYEN)

A. IMAGE ENHANCEMENT BASED ON IMPROVED RETINEX AND IMAGE FUSION

Under extremely low brightness and accompanied by a large amount of noise in underground mines, the quality of images captured by industrial cameras is poor, which reduces the detection accuracy greatly. Therefore, the improved Retinex algorithm is used in the image preprocessing stage to process images with low illumination and low contrast.

Retinex algorithm is an image enhancement algorithm based on the properties of the human eye visual system, which improves the brightness and color of an image through multi-scale processing of the image. Retinex image enhancement algorithm is classified into three categories, single-scale based Retinex algorithm (SSR), multi-scale based Retinex algorithm (MSR), and multi-scale Retinex algorithm with color restoration (MSRCR). However, all of them have defects after enhancement. Both SSR and MSR algorithms lose local color details during image enhancement, resulting in severe image distortion. MSRCR algorithm has the problem of haloing in regions with large differences in brightness during image enhancement. Therefore, an improved Retinex with a multi-image fusion algorithm is proposed. Three kinds of images after histogram equalization, adaptive luminance adjustment and improved Retinex processing are processed by channel fusion to retain their clear features, and finally the images are transferred to RGB color images to achieve image enhancement.

In the traditional Retinex algorithm, the scale factor σ of the Gaussian function determines the effect of image enhancement, however, this has obvious limitations. As shown in equation (1). G (x, y) is the Gaussian wrap around function expression.

$$G(x, y) = K^* \exp(-(x^2 + y^2)/(\sigma^2)) \tag{1}$$

where σ denotes the Gaussian surround scale; K is the normalization constant. To get the best image enhancement effect, we debug the optimal scale parameter of Gaussian surround function by experiment. According to the experimental results, when σ is relatively small, it can better maintain the edge feature information and expand the dynamic range but lacks in color recovery; when σ is relatively large, the color recovery is good, but the smaller dynamic range will cause the loss of detailed feature information. Therefore, it is determined that the parameter value of the Gaussian surround function is set to 80. $\sigma = 80$ can ensure good color recovery and maintain complete detailed feature information.

The improved Retinex algorithm uses Gaussian and bilateral filters as center surround functions. The incoming image to be processed is subjected to a convolution operation and the incident components are estimated separately. Then the estimated incoming and outgoing components are weighted and fused, and the fused incoming components can not only maintain the smoothness of the image and improve the image contrast but also better retain the image edge information to achieve the filtering effect. The construction process of the improve enhancement algorithm is shown in Fig. 5. The specific steps are as follows.

- Firstly, the image to be processed is converted to HSV color space and a threshold is set to adjust the luminance of its channel component V. Then it is converted to RGB color space.
- The processed image will be made into 3 copies; histogram equalization will be performed on the first copy and then median filtering will be performed on the processed image to eliminate the noise points present in the image.
- Automatic brightness adjustment is performed for the second copy. Firstly, the luminance value of the image component V is calculated, and the calculated luminance value is compared with a set threshold T value. The threshold T is set to 120, and the brightness value is higher than 120, the image is sufficiently bright and does not need to be processed; the brightness value is lower than 120, the image is insufficiently bright and needs to be subjected to a brightness enhancement operation.

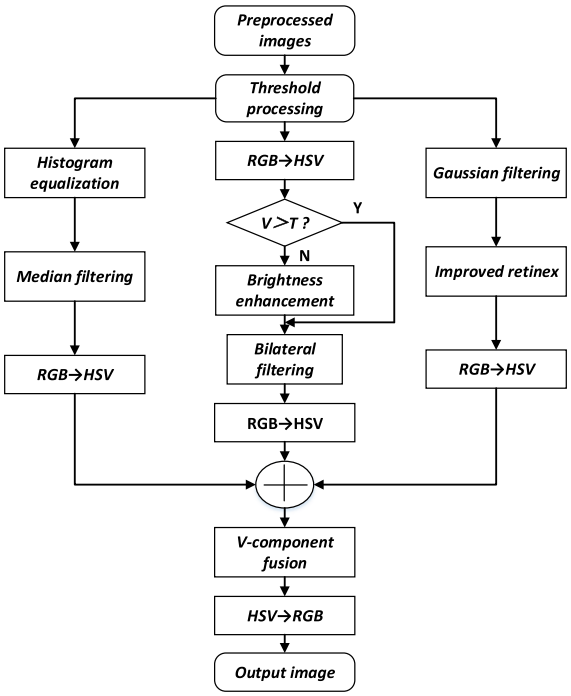

FIGURE 5. Flowchart of improved image enhancement algorithm.

Finally, the processed image is subjected to bilateral filtering to highlight the image edge information.

- The third copy is processed by the modified Retinex algorithm, which uses Gaussian filtering, bilateral filtering as its wraparound function to estimate the image illumination components separately, and finally outputs the reflectance map.
- The processed 3 copies of the image were transferred to HSV color space. Multi-image fusion is performed for its component V. The component values of the second image are used for both H and S components. The fused image is transferred from HSV to RGB color space, and the processed image is output.

A bilateral filter is a nonlinear filter with strong boundary protection, noise reduction, and smoothing capabilities, which is shown in equation (2). Similar to other filters, bilateral filtering also uses weighted averaging to represent the intensity of a pixel by weighting the luminance values of surrounding pixels. The difference is that bilateral filtering considers both weights. It considers not only the Euclidean distance of the pixels but also the difference in pixel paradigm thresholds.

$$g(x,y) = \frac{\sum_{(k,l)\in S(x,y)} f(k,l)w(x,y,k,l)}{\sum_{(k,l)\in S(x,y)} w(x,y,k,l)} \quad (2)$$

In Eq. (2): g(x, y) is the pixel gray value after bilateral filtering; f (k, l) is the gray value of the pixel point of the image to be processed; w(x, y, k, l) is the new gray value computed by using a weighted Gaussian function; S(x, y) is the range of gray values of the pixel point centered on point (x, y).

B. WEAR DETECTION BASED ON IMPROVED YOLOv8

The essence of conveyor belt wear condition detection is target detection. Target detection methods based on supervised learning can be categorized into two groups: two-stage target detection algorithms based on candidate regions and single-stage target detection algorithms based on regression. Two-stage target detection algorithms based on candidate regions generate candidate frames through a region proposal network (RPN) and then use a convolutional neural network for classification and localization regression. Such algorithms have higher accuracy, but due to the large number of candidate frames being slower, they are difficult to meet the real-time requirements. SPP-Net and R-CNN series are representative two-stage target detection algorithms. Regression-based single-stage target detection algorithms can directly regress the size, location and category of the target, greatly improving the detection speed. Representative ones are SSD (Single Shot MultiBox Detector) and YOLO (You Only Look Once).

1) YOLOv8 NETWORK ARCHITECTURE

YOLO algorithm has the advantages of fast speed, high accuracy, and strong generalization ability. Therefore, this study focuses on the most advanced YOLO detection model YOLOv8, whose network structure is shown in Fig.6.

The target detection network can be divided into three parts which are the Backbone feature extraction network part (Backbone), the Reinforcement feature extraction part (Neck), and the Prediction part (Head). The quality of the features extracted by the feature extraction network directly affects the prediction results of the prediction network. This means that if there are better features, there may be better prediction results. YOLOv8 provides a new SOTA model, that is CSPDarknet, as the backbone feature extraction network of YOLOv8, adopts the Cross Stage Partial Network (CSP) structure, which divides the network into two parts, each containing multiple residual blocks. The residual structure uses the idea of gradient diversion to effectively solve the problem of gradient explosion caused by network deepening. The deeper network also helps to extract deeper features and improve the accuracy of the network. Compared with Darknet, this structure can effectively reduce the parameters and computation of the model and improve the efficiency of feature extraction. Meanwhile, YOLOv8 draws on the idea of feature pyramid network (FPN). Multi-scale detection mechanism integrates three feature layers of different scales for target detection, which improves the network's ability to detect targets of different scales and increases the detection accuracy of small targets. Accuracy improves as the network depth deepens. However, in a limited number of computational units, the amount of computation likely increases with the depth of the network, which reduces the speed of target detection.

The backbone feature extraction network of YOLOv8 refers to the design idea of YOLOv7 ELAN. The C3 structure

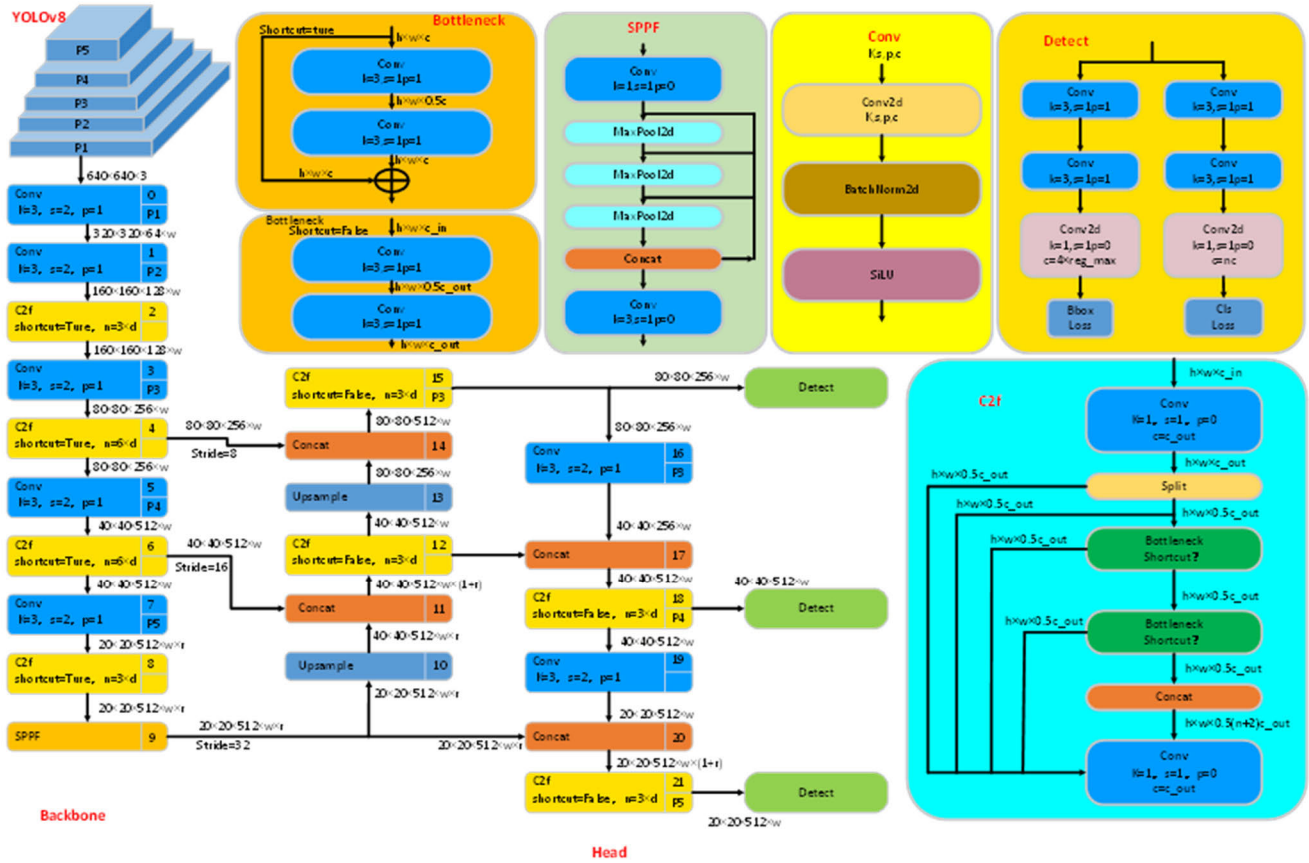

FIGURE 6. YOLOv8 network architecture.

of the backbone part and the neck part are all changed to the C2f structure with richer gradients, and the kernel of the first convolutional layer is changed from 6 × 6 to 3 × 3, which improves the detection speed of the model. The prediction part (head) is changed from coupled head to uncoupled head. The original anchor-based detection method is changed to anchor-free, which improves the generalization ability of the model and reduces the time cost of post-processing.

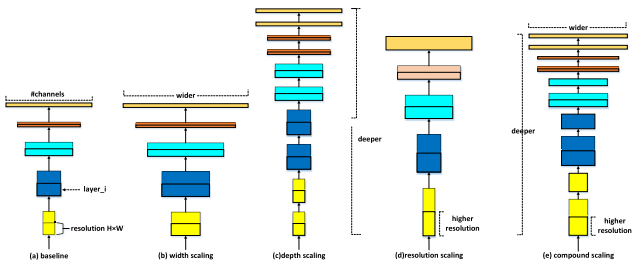

FIGURE 7. Structural scaling model.

2) SELECTION OF BACKBONE FEATURE NETWORK BASED ON EFFICIENTNET

For better extraction of features and information from image, enhanced and extended backbone feature extraction network methods are commonly used, which are shown in Fig.7. The base network is shown in Fig.(7-a). Width increasing method of the network is shown in Fig.(7-b), which increases the number of convolution kernels. More layer structures are used to increase the number of channels of the feature matrix, as well as increase the depth of the network, which is shown in Fig.(7-c). Resolution increasing method of the input network is shown in Fig.(7-d). However, this does not necessarily mean that network expansion can produce better results in terms of accuracy or speed. Increasing the depth of the network yields richer features and can be well applied to other tasks, but a network that is too deep can face problems of vanishing gradients and training difficulties. Increasing the width of the network results in more detailed features and is also easier to train, but it is often difficult to learn deeper features for networks with large widths and shallow depths. Increasing the image resolution of the input network can potentially yield higher fine-grained feature templates, but for very high input resolutions, the gain in accuracy is also reduced, and large-resolution images increase the amount of computation.

EfficientNet network utilises NAS (Neural Architecture Search) technology. This technique achieves a reasonable configuration of three parameters: network image input resolution, network depth, and channel width. The performance of the neural network is maximized under the limited resources,

and for the first time proposed and validated the model of the integrated scaling of the effect on the network as shown in Fig. (7-e).

EfficientNet draws on the idea of multi-objective optimization problems in mathematics and tries to use limited computational resources to improve the computational accuracy as much as possible by appropriately adjusting the depth, width, and resolution of the network's input images.

The algorithm consists of a convolutional network N of k convolutional layers, each of which can be represented as $Y_i = F_i(X_i)$.

The complete convolutional neural network structure can be represented as:

$$N = F_k \odot \cdots \odot F_2 \odot F_1(X_1) = \bigodot_{j=1,2,\ldots k} F_j(X_1) \quad (3)$$

The layers of convolutional neural networks commonly used in applications are divided into stages. Each stage of the convolutional layer has the same architecture, so the convolutional network can also be represented as:

$$\boldsymbol{N} = \bigodot_{i=1\ldots s} \boldsymbol{F}_i^{L_i}(X_{[H_i, W_i, C_i]}). \quad (4)$$

where: X_i denotes the input tensor, and $[H_i, W_i, C_i]$ denotes the shape of the input tensor X in the layer i, the Y_i denotes the output tensor, F_i denotes the operator operation, $F_i^{L_i}$ denotes that the F_i operation is repeated L_i times in the stage.

Conventional convolutional neural networks focus more on improving network performance by finding the optimal layer architecture. In contrast, EfficientNet neural networks simplify the design problem of new resource constraints by scaling $\boldsymbol{H_i}$, $\boldsymbol{W_i}$, $\boldsymbol{C_i}$ of the network without changing the predefined $\boldsymbol{F_i}$ in the baseline network. To further reduce the design space, it is restricted that all layers must be scaled uniformly at a constant rate. The mathematical model of the objective optimization can be expressed as:

$$Target : \frac{max}{d, w, r} Accuracy(N(d, w, r)) \quad (5)$$

$$s.t. N(d, w, r) = \bigodot_{i=1\ldots s} \hat{F}_i^{d \cdot \hat{L}i}(X_{\langle r \cdot \hat{H}_i, r \cdot \hat{w}_i, r \cdot \hat{c}_i \rangle}) \quad (6)$$

$$Memory(N) \leq target - memory \quad (7)$$

$$FLOPS(N) \leq target\_flops \quad (8)$$

where, w, d, and r are scale factors measuring the width, depth, and resolution of the network; $\hat{F}_i$, $\hat{L}_i$, $\hat{H}_i$, $\hat{W}_i$ and $\hat{C}_i$ are predefined parameters in the baseline network.

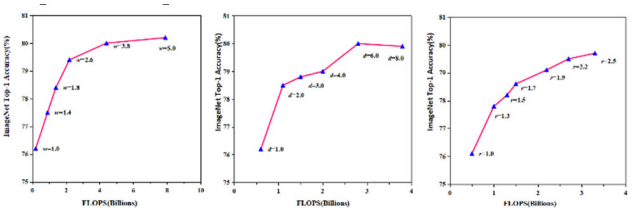

FIGURE 8. Limitations of single dimension expansion.

Therefore, for conveyor belts, large-resolution images that need to extract richer deep feature information require deeper networks. However, each dimension of the model scaling is not completely independent. It can cause the layer structure to lose balance and the network performance decline by expanding only a single dimension. EfficientNet proposes a hybrid scaling method, which uses a single mixing factor $\boldsymbol{\varphi}$ to uniformly scale width, depth, and resolution scaling. The specific calculation is as follows:

$$\textbf{depth} : \boldsymbol{d = \alpha^{\varphi}} \quad (9)$$

$$\boldsymbol{width} : \boldsymbol{w = \beta^{\varphi}} \quad (10)$$

$$\boldsymbol{resolution} : \boldsymbol{r = \gamma^{\varphi}} \quad (11)$$

$$\textbf{s.t.} \alpha \cdot \boldsymbol{\beta^2 \cdot \gamma^2} \approx 2 \quad (12)$$

$$\boldsymbol{\alpha \geq 1}, \quad \boldsymbol{\beta \geq 1}, \ \boldsymbol{\gamma \geq 1} \quad (13)$$

where: (α, β, γ) are constants determined using the NAS implementation. The theoretical computation amount is doubled as depth is doubled and will becomes 4 times as much as before as width and resolution are doubled. So, the total theoretical computation can be approximated by $(\alpha, \beta^2, \gamma^2)^{\varphi}$, which is equivalent to an increase of 2^{φ} times for any φ, after restricting $\alpha \cdot \beta^2 \cdot \gamma^2 \approx 2$.

The hybrid scaling method used by EfficientNet is scaled in two steps to determine the three parameters of α, β, γ.

Step 1: First, $\varphi = 1$ is fixed, and a twofold increase is assumed in available resources, then a mini-grid search is performed according to Eq. (5-13) to finally find the three best parameters (α, β, γ) for EfficientNet-B0.

Step 2: According to α, β and γ obtained in step 1, EfficientNet-B1 to EfficientNet-B7 are obtained based on EfficientNet-B0 using different methods.

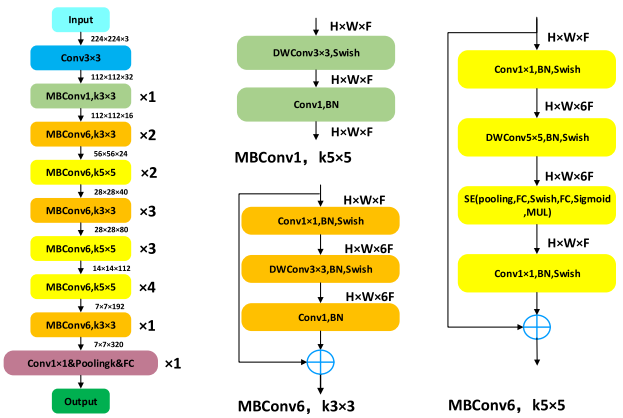

FIGURE 9. EfficientNet network architecture.

The structure of EfficientNet is shown in Fig. 9. The optimal EfficientNet is selected as the backbone feature extraction network of YOLOv8 according to the above steps to extract richer feature information for the prediction network of YOLOv8. After feature extraction by the backbone network of YOLOv8, the feature maps input to the prediction network are the combination of the original feature

maps after convolution downsampling and compression by a factor of 3, 4, and 5, which achieves favorable results in multi-scale target detection. This is the same as the EfficientNet network's down-sampling method for the input image. Therefore, the same downsampling method can realize the replacement of the YOLOv8 backbone feature extraction network.

3) NAM ATTENTION MODULE

To further enhance the accuracy of the improved network, the NAM attention mechanism is introduced at the channel where the backbone network extracts the three feature layers. Firstly, NAM serves as an efficient and lightweight attentional mechanism, which is employed to reduce less salient features and weights. Meanwhile, this module applies sparse weight penalties to make these weights computationally more efficient, enabling higher accuracy along with performance. Next, NAM utilizes the modular integration approach of CBAM, which redesigns the channel and spatial attention sub-modules. Hence, a NAM module is embedded at the end of each residual network. Noteworthy, For the channel attention submodule, a batch normalized (BN) scaling factor is used.

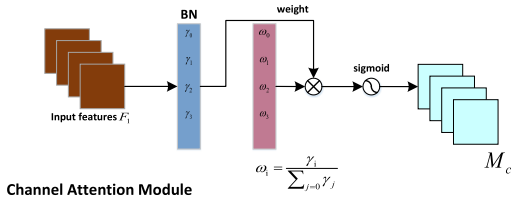

FIGURE 10. Channel attention mechanism.

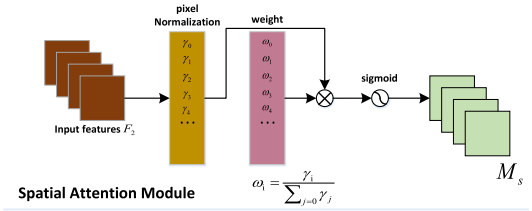

FIGURE 11. Spatial attention mechanism.

The channel attention sub-module is shown in Fig. 10, and its mathematical expression is given in Eq. (14). which represents the output features obtained at the end, and γ is the scaling factor for each channel. This gives the weights for each channel. If the same normalization method is employed for each pixel in the space, the weights for spatial attention can be obtained. The pixel normalization method is shown in Fig. 11 and its mathematical expression is given as Eq. (15). To suppress unimportant features, a regularization term is added to the loss function as in Eq. (16).

$$\boldsymbol{M_c} = \mathbf{sigmoid}(\boldsymbol{W_\gamma}(\mathbf{BN}(\boldsymbol{F_1}))) \tag{14}$$

$$\boldsymbol{M_s} = \mathbf{sigmoid}(\boldsymbol{W_\gamma}(\mathbf{BN_s}(\mathbf{F_2}))) \tag{15}$$

$$\mathbf{Loss} = \sum_{(x,y)} \boldsymbol{l}(\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{W}), \boldsymbol{y}) + \boldsymbol{p}\boldsymbol{\Sigma}\boldsymbol{g}(\boldsymbol{\gamma}) + \boldsymbol{p}\boldsymbol{\Sigma}\boldsymbol{g}(\boldsymbol{\lambda}) \tag{16}$$

4) RETINEX-YOLOv8-EFFICIENTNET-NAM DETECTION MODEL(RYEN)

The structure of Retinex-YOLOv8-EfficientNet- NAM (RYEN) is shown in Fig.12. Improved Retinex with a multi-image fusion algorithm is used to preprocess the captured images. YOLOv8 target detection algorithm is used as the basis. Then, the original backbone feature extraction network CSPDarknet of YOLOv8 is replaced by EfficientNet, which makes the original huge network structure lighter and optimizes the scaling of the network without changing the layer structure. This ensures the original detection accuracy of YOLOv8. Furthermore, incorporating the NAM lightweight attention module into the three feature layer channels of the predictive network enhances the extraction of backbone features, resulting in improved network detection accuracy.

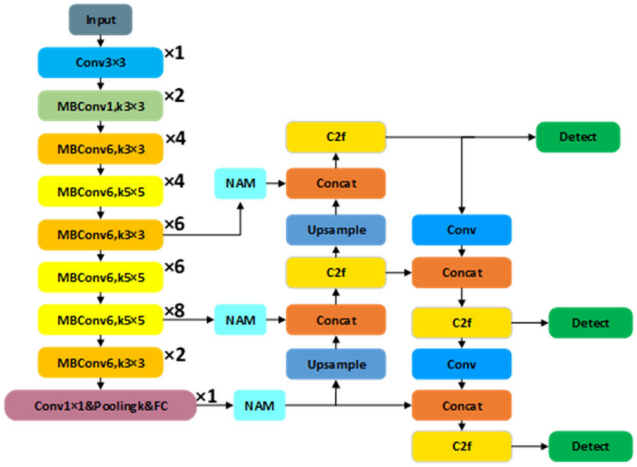

FIGURE 12. RYEN network structure diagram.

IV. EXPERIMENTS AND DISCUSSIONS

A. COMPARISON OF ALGORITHMS BEFORE AND AFTER IMAGE ENHANCEMENT

The data set used in this experiment was collected from a mine plant of Jizhong Energy Fengfeng Group. Hikvision MV-CA013-20GC industrial camera was used to capture the dataset. The real camera is shown in Figure 13. Before shooting, the camera was fixed on the unloading part of the conveyor below the mine, as shown in Figure 14 in the way and location of erection. It was set to manual shooting mode to adjust the Angle and replenish the light source.

Fewer data resources in the process of conveyor belt wear recognition can cause overfitting of the deep network model. Fewer data samples can also result in less robust network models and weaker generalization. Therefore, random cropping, rotation, scaling, brightness adjustment, and adding pretzel noise are used to expand the dataset. After expansion, there are a total of 1831 images in the dataset. The training set, validation set, and test set are divided in a ratio of 7:2:1. LabeImg software is used to annotate the wear parts of the

FIGURE 13. Spatial attention mechanism.

FIGURE 14. Physical picture of the position of the frame camera.

conveyor belt in the image, and the type of belt wear degree and coordinate information are saved into an XML file.

The image processed by the improved Retinex with multiple image fusion algorithm is shown in Fig.15. Reliance solely on camera acquisition without any processing methods in underground mines can result in very low image brightness. The captured images are very dark and blurry, almost invisible. Therefore, they cannot meet the requirements of any detection model. The image appears brighter after undergoing the original Retinex processing. However, the presence of noise and other distortions in the mine results in reduced clarity of the image. Additionally, the worn texture of the belt is not clearly visible, and the image is noticeably distorted. The detection effect of the detection model is also greatly affected. For the complex environment of the mine, it is difficult for the common image enhancement algorithms to meet the requirements. Improved Retinex multi-image fusion algorithm, first the original Retinex algorithm is improved and then the processed image with the improved algorithm is channel fused with the multi-image. The processed image is greatly improved in terms of brightness, sharpness and contrast.

To verify its effectiveness, the original image, the Retinex processed image, and the image processed by the improved image enhancement algorithm are detected on the conveyor test bed. The results are shown in Fig. 16.

The image captured without any means of processing is too low in brightness, resulting in the detection model is unable

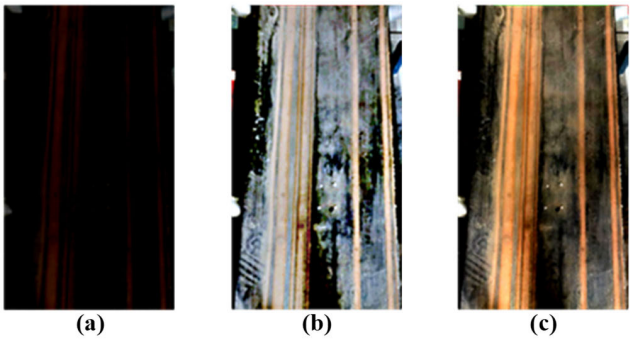

FIGURE 15. Image enhancement comparison (a) Original image (b) Retinex processed image (c) Improved Retinex processed images.

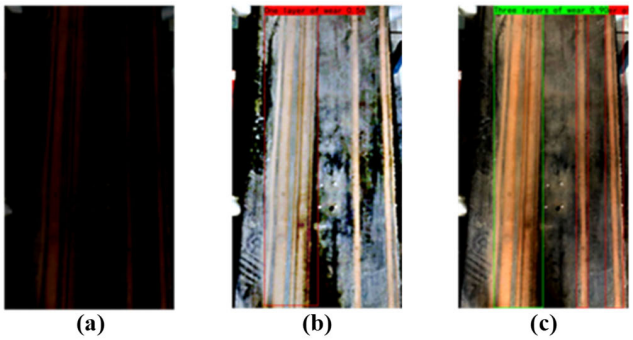

FIGURE 16. Comparison of the effect of image enhancement algorithms after applying to the algorithm. (a) Original image (b) Retinex processed image (c) Improved Retinex processed images.

to detect the belt wear state. There are some problems by Retinex image processing, such as wrong wear classification, low recognition accuracy and missed detection. In contrast, the improved Retinex algorithm detects clear images, recognises wear categories accurately, has high accuracy and is well suited to the detection model.

B. COMPARATIVE EXPERIMENT

The hardware platform configuration for this experimental model training, CPU for AMD Ryzen 7 6800H; GPU for NVIDIA GeForce RTX 3070 Ti; experimental environment for Windows 11; Pytorch1.13.0; CUDA11.7; Python3.9.

In this paper, the performance of the model is evaluated in terms of both detection accuracy and detection speed. For the detection performance, mean average precision (mAP), computation (GFLOPs), number of parameters (Params), and speed (FPS) are used as evaluation metrics. The precision of calculating mAP needs to calculate average precision (AP) first, and then average the AP values of different classes. The AP value is obtained by calculating the average of the precision values for each recall value, as shown in Eq. (17-20).

$$Precision = \frac{TP}{TP + FP} \tag{17}$$

$$Recall = \frac{TP}{TP + FN} \tag{18}$$

TABLE 1. Performance of different backbone feature networks on the dataset.

Model-Name	mAP/%	FPS	Wear one layer	Wear two layers	Wear three layers	Wear four layers	Wear five layers
EfficientNet-B0-Yolov8	88.45	81.81	83	83	86	91	100
EfficientNet-B1-Yolov8	90.32	62.67	87	85	86	92	100
EfficientNet-B2-Yolov8	93.84	60.28	89	89	89	98	100
EfficientNet-B3-Yolov8	94.59	59.15	88	90	91	99	100
EfficientNet-B4-Yolov8	98.57	61.11	96	98	100	100	100
Yolov8	92.13	52.94	89	91	94	100	100

where TP is a positive category judged as positive, FP is a negative category judged as positive, and FN is a positive category judged as negative.

$$AP = \int_0^1 \rho(r)d(r) \tag{19}$$

$$mAP = \frac{1}{K}\sum_{i=1}^{K} AP_i \tag{20}$$

where, ρ is the precision, which indicates the proportion of correct results identified by the model among all results identified; r is the recall, which indicates the proportion of correct results identified by the model among the results that need to be identified in the dataset; and K is the number of categories in the sample.

1) BACKBONE FEATURE EXTRACTION NETWORK TESTING

To select the backbone feature network structure with optimal performance, EfficientNet B0-B7 is tested as the backbone of Yolov8, respectively. The test performance is shown in Table 1. The algorithm achieves the best speed-accuracy trade-off on the conveyor belt wear dataset with an AP rate of 98.57% at 61.11 FPS and 88.45% at 81.81 FPS. Within the constraints of the experimental environment, the requirements on the hardware devices increase as the amount of computation increases, and the problem of graphic memory overflow occurs when experimenting with EfficientNet-B5. However, the trend from Table 1 shows that as the coefficient φ increases, the model is further extended, and the prediction accuracy of the neural network is further improved. However, the increase in computation inevitably reduces the detection speed. Therefore, EfficientNet-B4 with better performance is chosen, which achieves an AP rate of 98.57% at 61.11 FPS.

2) COMPARISON OF ALGORITHM PERFORMANCE WITH NAM ATTENTION MODULE

Under the same experimental conditions, different attention modules are added to YOLOv8 algorithm to compare the effects of different attention modules on the performance of the algorithm, the results are shown in Table 2 and Fig.17. Among them, CAM is a channel attention module with a compression rate of 16, and SE is a lightweight attention module. As seen from table 2, the addition of a single channel attention module CAM improved the detection accuracy by 1.1% compared to YOLOv8. The improvement effect is not significant.

TABLE 2. Performance comparison after introduction of attention module.

Models	Params	GFLOPs	mAP50	FPS
YOLOv8 -NAM	38.3M	110.4G	97.27%	62.1
YOLOv8 -CBAM	38.9M	111.2G	95.01%	51.5
YOLOv8 -SE	36.2M	98.3G	94.09%	59.3
YOLOv8 -CAM	38.7M	108.1G	93.25%	58.2
YOLOv8	37.6M	110.4G	92.13%	52.9

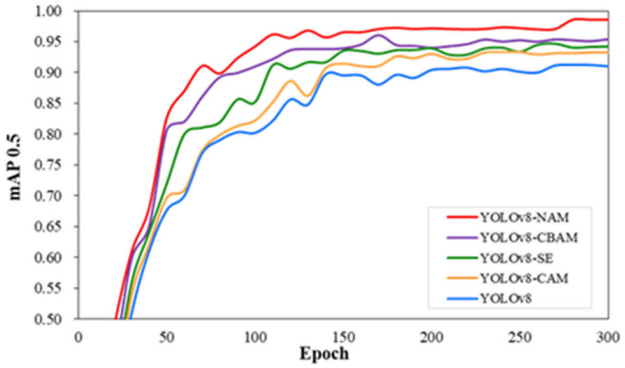

FIGURE 17. Results of different attention module training.

For the CBAM attention module with channel-space synergy, the number of parameters is increased by 1.3M, the computation is increased by 0.8G, and the detection accuracy is increased by 2.88% compared to the original algorithm. The CBAM mechanism ignores channel-space interactions and thus loses cross-latitude information. Compared with CBAM, the number of parameters of the NAM attention module is reduced by 0.6 G, the amount of computation of NAM is reduced by 0.8 G, and the detection accuracy of NAM is improved by 2.26%. The detection accuracy of NAM is improved by 5.14% compared to the original algorithm with a frame rate of 62.1 per second. The experiments show that NAM attention module highlights salient features by utilizing the variance measure of the training model weights, resulting in more accurate detection results.

3) COMPARISON OF RYEN ALGORITHM AND YOLOv8 PERFORMANCE

Performance comparison of RYEN algorithm and the original YOLOv8 on the dataset is shown in Fig.16 and Table 3. As shown in Fig. 18, both algorithms are better able to

TABLE 3. PERformance comparison of ryen algorithm and original YOLOV8.

species	YOLOv8				RYEN algorithm			
	precision/%	recall/%	F1 score/%	AP/%	precision/%	recall/%	F1 score/%	AP/%
Wear one layer	94.23	82.61	91.00	91.60	98.18	83.08	90.00	94.40
Wear two layers	92.15	83.67	94.00	94.38	94.26	87.50	93.00	98.07
Wear three layers	93.54	90.24	91.00	92.04	100.00	100.00	100.00	100.00
Wear four layers	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Wear five layers	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
average	95.98	91.30	95.20	95.60	98.49	94.12	96.60	98.49

TABLE 4. Performance comparison of various algorithms.

Model-Name	Size	Params	GFLOPs	mAP	FPS	CPU	RAM	CUDA	Identify the time/s
Faster R-CNN	640	137.099M	402.181G	89.57	13.84	41	23	98	0.08221
SSD	640	26.285M	281.966G	90.55	82.17	21	29	84	0.01216
YOLOv5	640	47.057M	115.918G	90.91	55.61	80	25	75	0.03098
YOLOv7	640	37.620M	106.472G	90.13	48.95	51	23	74	0.03034
YOLO8	640	37.641M	110.374G	92.13	52.94	34	28	68	0.03309
RYEN algorithm	640	29.167M	87.817G	98.57	66.11	31	21	62	0.02512

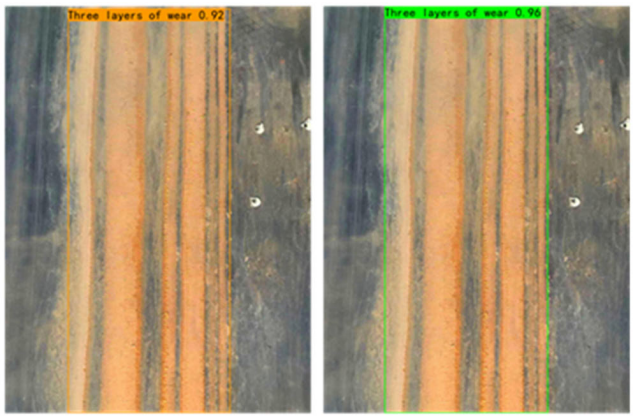

FIGURE 18. Comparison of experimental effect of YOLOv8 before and after improvement. (a) YOLOv8 Detection Effect (b) RYEN algorithm detection effect.

achieve the identification of various predetermined conveyor belt wear areas and the determination of the degree of wear. It is obtained from Table 3 that, compared with the original YOLOV8 algorithm, RYEN algorithm has higher accuracy and recall rate for the belt with three layers of wear. This is very important because, according to the data obtained after our mechanical performance tests, a wear level of three layers means that the performance of the belt reaches a level close to substandard. The maximum tensile strength that a worn three-ply belt can withstand already exceeds the national standard and poses a safety hazard. As seen from Fig. 18, the accuracy of the RYEN algorithm is slightly higher than the original algorithm, which is consistent with the results of table 3. Furthermore, the RYEN algorithm has higher recognition accuracy for detecting three layers of wear.

4) PERFORMANCE COMPARISON OF DIFFERENT ALGORITHMS

The performances of several algorithms are tested simultaneously on the conveyor belt wear dataset, including Faster-RCNN, SSD, YOLOv5, YOLOv7, YOLOv8, and RYEN algorithm. The network sets are to be trained with 300 iterations, Adam's initial learning rate is 0.001, the weight decay coefficient is set to 0.0005, and the learning rate momentum is 0.937. The performance results of various algorithms are shown in Table 4.

Seen in Table 4, the detection speed of Faster R-CNN, is 13.84 FPS, which is the slowest among all tested algorithms. RPN network that takes the extraction of candidate regions replaces the previous selective search, which greatly reduces the amount of computation. However, compared with YOLO algorithm or SSD algorithm which needn't generate candidate frames (region proposal) and is directly based on regression prediction, computational volume of RPN network is still large, and it is difficult to meet the real-time demand. In contrast, the SSD algorithm based on the regression mechanism also predicts multi-scale feature maps, which greatly improves the ability to detect large and small targets simultaneously and achieves an accuracy of 90.55% in the dataset. On the other hand, YOLO algorithm has very good results in terms of accuracy and detection speed. Based on YOLOv8,

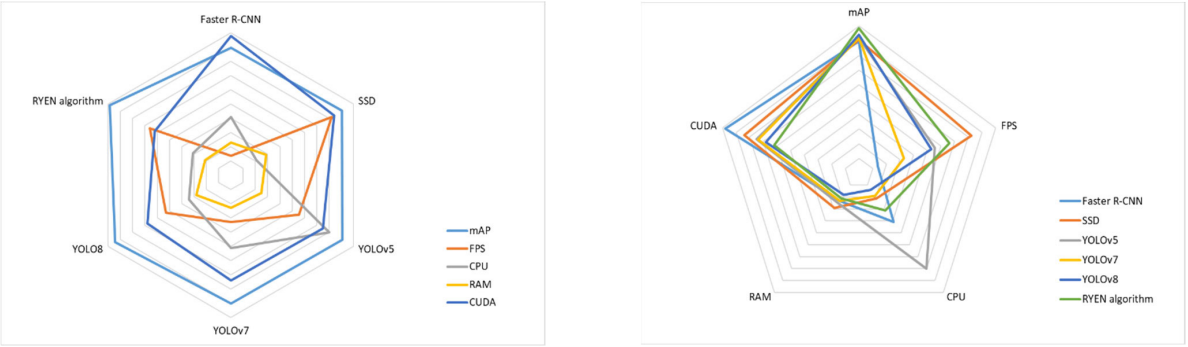

FIGURE 19. Utilization of different hardware.

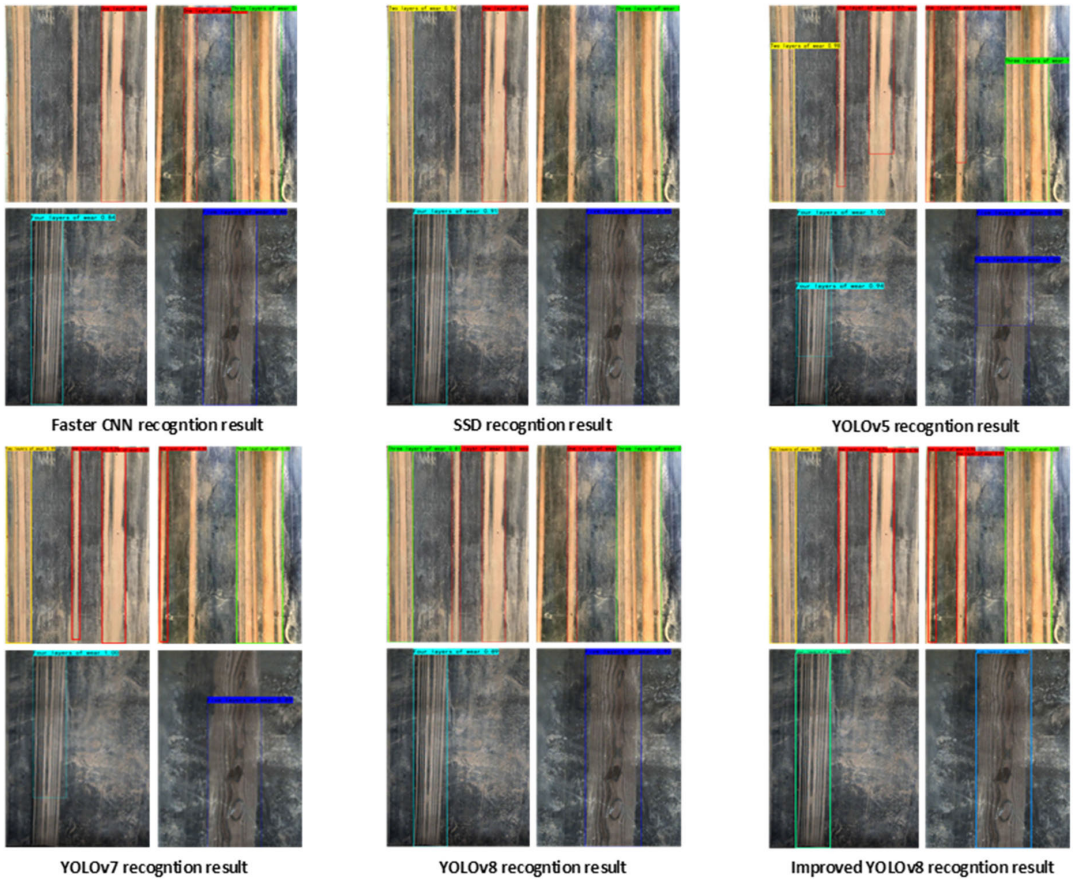

FIGURE 20. Recognition effect of different algorithm.

RYEN algorithm proposed in this paper has accuracy rate of 98.57 % and detection speed of 66.11 FPS.

To evaluate the performance of the algorithms more comprehensively, in addition to considering the algorithm accuracy and detection speed aspects, the hardware usage of different algorithms is also compared. The performance results are shown in Fig.19, which is more instructive for the operators in the actual working conditions.

As can be seen from Fig.19, the GPU and RAM usage of RYEN algorithm (here the backbone feature extraction network is EfficientNet-B4) is less compared to the original YOLOv8, and the CPU usage of both is the same. In comparison, the huge computational volume of Faster R-CNN results in it taking up a large number of GPUs and CPUs. In addition, it is more obvious that the RYEN algorithm based on the EfficientNet feature extraction network proposed in this paper has higher timeliness while guaranteeing the accuracy of the measurement.

Four belt wear images with different wear levels are selected to be predicted with different algorithms. The prediction results are shown in Fig.20. Faster-R-CNN and SSD have the problem of not detecting worn regions.

TABLE 5. Comparison of the accuracy of different detection models and manual detection.

Abrasion category	Artificial		Faster RCNN		SSD		YOLOv5		YOLOv7		YOLOv8		RYEN algorithm	
	Adr	Cdr	Adr	Cdr	Adr	Cdr	Adr	Cdr	Adr	Cdr	Adr	Cdr	Adr	Cdr
One layer of wear	57.89	63.64	68.84	53.85	52.63	70.00	63.16	58.33	78.95	66.67	80.13	79.32	89.47	88.24
Two layers of wear	52.13	38.46	52.26	53.85	56.00	42.86	44.86	45.45	64.38	58.25	73.56	78.61	92.13	86.96
Three layers of wear	48.15	46.15	62.96	35.29	59.26	68.75	44.44	50.00	70.37	73.69	78.64	80.74	92.59	91.30
Four layers of wear	66.67	37.50	83.33	50.00	75.00	66.67	83.33	60.00	91.63	72.73	91.65	86.91	91.67	90.91
Five layers of wear	70.06	41.67	82.35	43.75	76.47	61.54	82.35	57.14	88.24	73.33	88.71	88.74	94.12	93.75
average	58.98	45.48	69.95	47.35	63.87	61.96	63.63	54.18	78.71	68.93	82.538	81.395	92.00	90.23

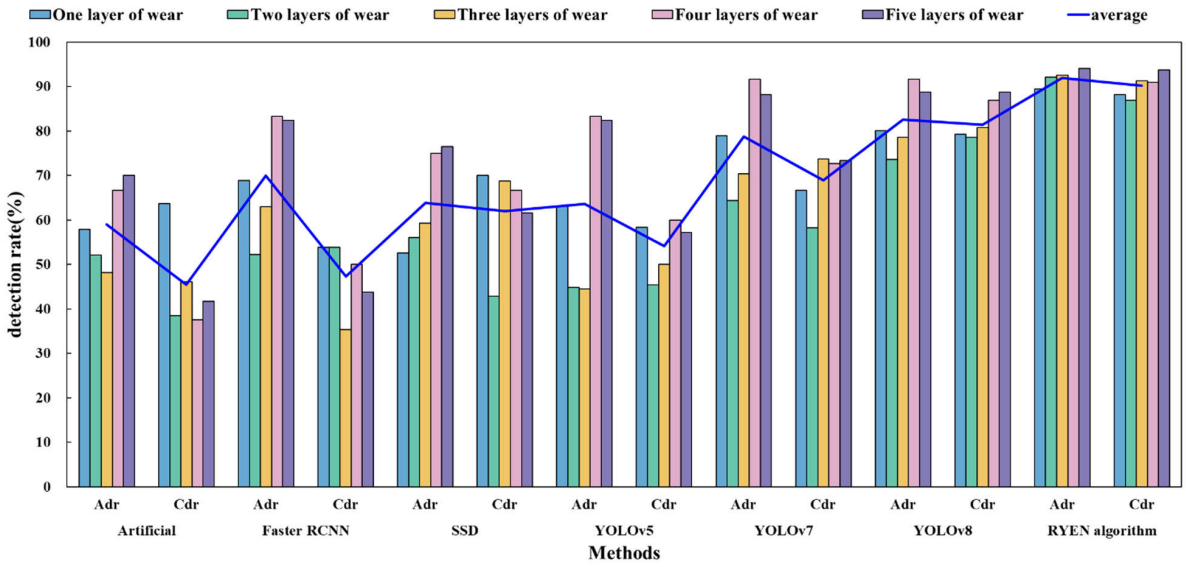

FIGURE 21. Comparison of accuracy data between different detection models and manual detection.

YOLOv5 and YOLOv7 also have the problem of inaccurate frame prediction. Compared with other algorithms, YOLOv8 has more outstanding detection ability, but its large structure leads to slower detection speed. The RYEN algorithm detects worn regions with higher confidence and accuracy than the original YOLOv8. The RYEN algorithm performs better in terms of accuracy. The RYEN algorithm performs better in terms of accuracy.

Compared to the original YOLOv8, the RYEN algorithm has an average test accuracy of 98.57% on the dataset. The accuracy of the improved algorithm has been improved, especially in shallow wear detection. Moreover, the improved algorithm has significantly improved the prediction speed, increasing the number of frames from 52.94 FPS to 61.11 FPS.

5) COMPARISON OF DIFFERENT NETWORK DETECTION MODELS WITH MANUAL DETECTION

Different network detection models are compared with manual detection by experiments. Experiments are conducted by placing 100 pre-prepared conveyor belts with different levels of wear on the belt conveyor. Among them, 19 wear points in one layer, 25 wear points in two layers, 27 wear points in three layers, 12 wear points in four layers, and 17 wear points in five layers. The average detection rate (Adr) and the correct detection rate (Cdr) are calculated by counting the number of wear and tear and the number of correct identifications, respectively, for different detection models and manual detection. The comparative data results are shown in Fig.19 and Table 5. The average detection rate was 58.98 per cent and the correct detection rate was 45.48 per cent, depending only on the subjective judgement of the inspectors. The reason is that the inspector's position for observing the conveyor belt is not the optimal observation position, and the angle of vision and the running speed of the conveyor belt can lead to missed detection and incorrect detection of the degree of wear. So, Small target wear points cannot be detected. The average detection rate and correct detection rate of different network inspection models are higher than manual inspection.

C. VALIDATION OF MODEL GENERALIZATION CAPABILITIES

The generalization ability of the model is the key to the success of the model. To verify the generalization ability of

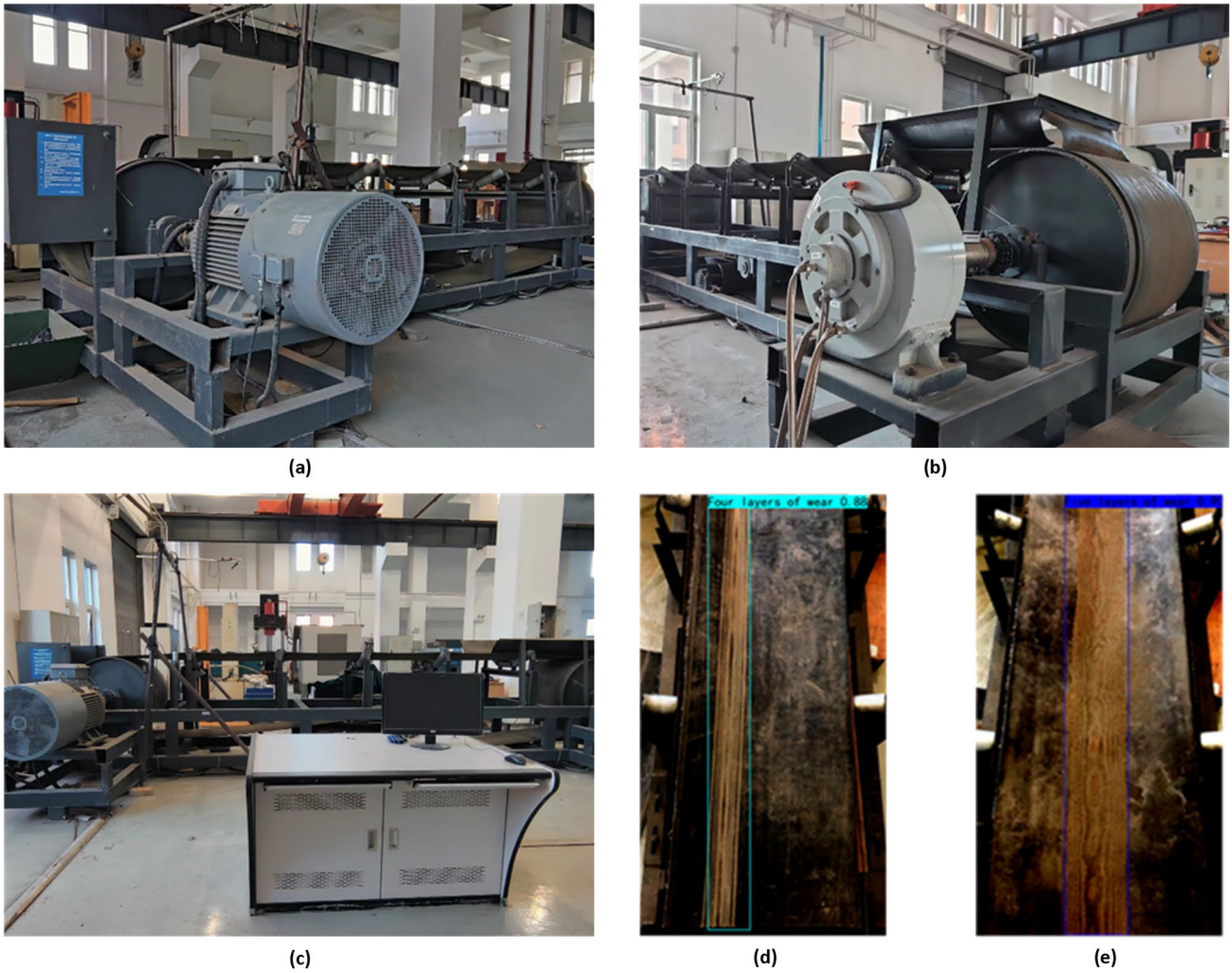

FIGURE 22. Field test equipment and experimental results. (a) (b) (c) Conveyor test rig at Jizhong Energy Fengfeng Group's mine plant. (d) (e) Belt wear test field test results.

the proposed model, a laboratory simulation environment test is conducted.

Since it was impossible to apply the belts with wear requirements to the conveyor, we have chosen to affix the broken belts to a normally operating conveyor belt test bed, which is shown in Fig. 22 (a) (b) (c). A Hikvision MV-CA013-20GC industrial camera is mounted directly above the unloaded belt. The camera frame rate is 90 fps, and the maximum resolution is 1280 × 1024. The width of the conveyor belt is 1.4 m, and the rated running speed is 3.0 m/s. As shown in Fig.22(d) (e), the algorithm accurately identifies belt conveyor belts with four and five layers of wear, with accuracy rates of 88% and 90%, and real-time frame rates of 70.62 FPS and 74.28 FPS, respectively, meeting the actual needs of the project.

V. CONCLUSION

In this paper, a new detection method RYEN algorithm of the wear state detection of mining conveyor belts based on deep learning and machine vision is proposed. The degree of belt wear is reclassified and redefined with the support of the mechanical properties data of the worn belt and the wear texture features, to establish a new dedicated dataset for conveyor belt wear detection.To solve the problems of low brightness and high noise under the mine with complex conditions, the acquired images are preprocessed based on improved Retinex with multi-image fusion.Then,based on the YOLOv8 algorithm, EfficientNet is used to replace Darknet53 as the backbone feature extraction network, comprehensively considering the balance of network scaling network depth, width, and image resolution, and improving the accuracy of the algorithm under limited computing resources. Moreover, a lightweight attention module NAM is introduced, which increases the accuracy of belt wear detection while improving the detection speed. RYEN algorithm is verified with better detection accuracy and real-time performance through field experiments. The main contributions of this paper are summarized as follows.

(1)Aiming at the daily fatigue wear of conveyor belts as the main concern, the degree of wear of different wear states of belts is redefined and classified by analyzing the data of mechanical property tests and wear texture characteristics. A more detailed classification of belt wear states can more accurately monitor the operating condition of the conveyor belt and make the belt utilization rate maximized.

(2) Aiming at the problems of low brightness, high noise, and complex working conditions under the mine, improved Retinex with a multi-image fusion algorithm is used to preprocess the captured images.

(3) Considering the balance of network scaling, network depth, width, and image resolution, EfficientNet is used to replace Darknet53 of YOLOv8 as the backbone feature extraction network to improve detection accuracy under limited computing resources. Moreover, a lightweight attention module NAM is added to the three channels of the backbone feature extraction network of YOLOv8 to improve the detection speed without reducing detection accuracy.

(4) RYEN algorithm results in a double improvement in the detection accuracy and detection speed. The tested detection accuracy on the conveyor belt wear dataset is as high as 98.57%, and the fastest tested detection speed is 66 FPS.

From the test platform, it can be seen that our proposed model has high generalization ability. Comparing the detcction accuracy and speed of the algorithms, the hardware usage is also considered when the algorithms are running to provide a reference for engineers to choose the algorithms. In the next work, the focus of research will be on how to make the model more lightweight. Meanwhile, we will further extend the belt-wear dataset to improve the robustness and generalization ability of the neural network.

REFERENCES

[1] M. Zhang, H. Shi, Y. Yu, and M. Zhou, "A computer vision based conveyor deviation detection system," *Appl. Sci.*, vol. 10, no. 7, p. 2402, Apr. 2020.

[2] T. T. Kwo, "A theory of conveyors," *Manage. Sci.*, vol. 5, no. 1, pp. 51–71, Oct. 1958.

[3] Y. Chunxia and C. Si, "A new detection device of belt longitudinal rip based on pressure monitoring," in *Proc. Int. Conf. Electron., Commun. Control (ICECC)*, Sep. 2011, pp. 60–63.

[4] G. Fedorko, V. Molnar, A. Grincova, M. Dovica, T. Toth, N. Husakova, V. Taraba, and M. Kelemen, "Failure analysis of irreversible changes in the construction of rubber–textile conveyor belt damaged by sharp-edge material impact," *Eng. Failure Anal.*, vol. 39, pp. 135–148, Apr. 2014.

[5] G. Fedorko, V. Molnár, Ž. Ferková, P. Peterka, J. Krešák, and M. Tomašková, "Possibilities of failure analysis for steel cord conveyor belts using knowledge obtained from non-destructive testing of steel ropes," *Eng. Failure Anal.*, vol. 67, pp. 33–45, Sep. 2016.

[6] P. Dąbek, A. Wróblewski, J. Wodecki, P. Bortnowski, M. Ozdoba, R. Król, and R. Zimroz, "Application of the methods of monitoring and detecting the belt mistracking in laboratory conditions," *Appl. Sci.*, vol. 13, no. 4, p. 2111, Feb. 2023.

[7] R. Volodymyr, "Method and apparatus for eddy current-based quality inspection of dry electrode structure," U.S. Patent 7 355 395, Apr. 2008.

[8] T. Qiao, X. Lu, and L. Yan, "Research on the signal feature extraction method in steel-cord conveyor belt with metal magnetic memory testing," *Adv. Sci. Lett.*, vol. 11, no. 1, pp. 489–492, May 2012.

[9] Y. Guan, J. Zhang, Y. Shang, M. Wu, and X. Liu, "Embedded sensor of forecast conveyer belt breaks," in *Proc. 5th Int. Conf. Fuzzy Syst. Knowl. Discovery*, vol. 5, Oct. 2008.

[10] Y. Pang, "A novel embedded conductive detection system for intelligent conveyor belt monitoring," in *Proc. IEEE Int. Conf. Service Operations Logistics, Informat.*, Nov. 2006.

[11] G. Wang, X. Li, and L. Yang, "Dynamic coal quantity detection and classification of permanent magnet direct drive belt conveyor based on machine vision and deep learning," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 35, no. 11, Sep. 2021, Art. no. 2152017.

[12] H. I. Bozma and H. Yalçın, "Visual processing and classification of items on a moving conveyor: A selective perception approach," *Robot. Comput.-Integr. Manuf.*, vol. 18, no. 2, pp. 125–133, Apr. 2002.

[13] J. Tessier, C. Duchesne, and G. Bartolacci, "A machine vision approach to on-line estimation of run-of-mine ore composition on conveyor belts," *Minerals Eng.*, vol. 20, no. 12, pp. 1129–1144, Oct. 2007.

[14] M. Kontny, "Machine vision methods for estimation of size distribution of aggregate transported on conveyor belts," *Vibroengineering Proc.*, vol. 13, pp. 296–300, Sep. 2017.

[15] M. Kistner, G. T. Jemwa, and C. Aldrich, "Monitoring of mineral processing systems by using textural image analysis," *Minerals Eng.*, vol. 52, pp. 169–177, Oct. 2013.

[16] C. Aldrich, "Online analysis of coal on a conveyor belt by use of machine vision and kernel methods," *Int. J. Coal Preparation Utilization*, vol. 30, pp. 331–348, Nov. 2010.

[17] Z. Zhang, Y. Liu, Q. Hu, Z. Zhang, L. Wang, X. Liu, and X. Xia, "Multi-information online detection of coal quality based on machine vision," *Powder Technol.*, vol. 374, pp. 250–262, Sep. 2020.

[18] Y. D. Zhao and M. F. Sun, "Image processing and recognition system based on Davinci technology for coal and gangue," *Appl. Mech. Mater.*, vols. 130–134, pp. 2107–2110, Oct. 2011.

[19] M.-H. Zhao, S. Ma, and D.-X. Zhao, "Image processing based on gray information in sorting system of coal gangue," in *Proc. 10th Int. Conf. Intell. Human-Machine Syst. Cybern. (IHMSC)*, vol. 2, Aug. 2018, pp. 81–83.

[20] T. Andersson, M. J. Thurley, and J. E. Carlson, "A machine vision system for estimation of size distributions by weight of limestone particles," *Minerals Eng.*, vol. 25, no. 1, pp. 38–46, Jan. 2012.

[21] B. Ryszard, R. Zimroz, L. Jurdziak, M. Hardygora, and W. Kawalec, "Conveyor belt condition evaluation via non-destructive testing techniques," in *Proc. 22nd MPES Conf.*, Dresden, Germany, Oct. 2014, pp. 1119–1126.

[22] X. Cao, X. Zhang, Z. Zhou, J. Fei, G. Zhang, and W. Jiang, "Research on the monitoring system of belt conveyor based on suspension inspection robot," in *Proc. IEEE Int. Conf. Real-Time Comput. Robot. (RCAR)*, Aug. 2018, pp. 657–661.

[23] S. Nashat, A. Abdullah, S. Aramvith, and M. Z. Abdullah, "Support vector machine approach to real-time inspection of biscuits on moving conveyor belt," *Comput. Electron. Agricult.*, vol. 75, no. 1, pp. 147–158, Jan. 2011.

[24] M. Zhang, K. Jiang, Y. Cao, M. Li, N. Hao, and Y. Zhang, "A deep learning-based method for deviation status detection in intelligent conveyor belt system," *J. Cleaner Prod.*, vol. 363, Aug. 2022, Art. no. 132575.

[25] J. Chamorro, L. Vallejo, C. Maynard, S. Guevara, J. A. Solorio, N. Soto, K. V. Singh, U. Bhate, J. Garcia, and B. Newell, "Health monitoring of a conveyor belt system using machine vision and real-time sensor data," *CIRP J. Manuf. Sci. Technol.*, vol. 38, pp. 38–50, Aug. 2022.

[26] X.-L. Hao and H. Liang, "A multi-class support vector machine real-time detection system for surface damage of conveyor belts based on visual saliency," *Measurement*, vol. 146, pp. 125–132, Nov. 2019.

[27] Z. Lv, X. Zhang, J. Hu, and K. Lin, "Visual detection method based on line lasers for the detection of longitudinal tears in conveyor belts," *Measurement*, vol. 183, Oct. 2021, Art. no. 109800.

[28] T. Qiao, W. Liu, Y. Pang, and G. Yan, "Research on visible light and infrared vision real-time detection system for conveyor belt longitudinal tear," *IET Sci., Meas. Technol.*, vol. 10, no. 6, pp. 577–584, Sep. 2016.

[29] C. Hou, T. Qiao, H. Zhang, Y. Pang, and X. Xiong, "Multispectral visual detection method for conveyor belt longitudinal tear," *Measurement*, vol. 143, pp. 246–257, Sep. 2019.

[30] C. Hou, T. Qiao, M. Qiao, X. Xiong, Y. Yang, and H. Zhang, "Research on audio-visual detection method for conveyor belt longitudinal tear," *IEEE Access*, vol. 7, pp. 120202–120213, 2019.

[31] X. Guo, X. Liu, H. Zhou, R. Stanislawski, G. Królczyk, and Z. Li, "Belt tear detection for coal mining conveyors," *Micromachines*, vol. 13, no. 3, p. 449, Mar. 2022.

[32] X. Guo, X. Liu, G. Królczyk, M. Sulowicz, A. Glowacz, P. Gardoni, and Z. Li, "Damage detection for conveyor belt surface based on conditional cycle generative adversarial network," *Sensors*, vol. 22, no. 9, p. 3485, May 2022.
[33] M. Liu, Q. Zhu, Y. Yin, Y. Fan, Z. Su, and S. Zhang, "Damage detection method of mining conveyor belt based on deep learning," *IEEE Sensors J.*, vol. 22, no. 11, pp. 10870–10879, Jun. 2022.
[34] Y. Wang, Y. Wang, and L. Dang, "Video detection of foreign objects on the surface of belt conveyor underground coal mine based on improved SSD," *J. Ambient Intell. Humanized Comput.*, vol. 14, no. 5, pp. 5507–5516, Sep. 2020.

LIJIE YANG was born in Hebei, China, in 1980. She received the bachelor's degree in mechanical design, manufacturing and automation from the Hebei Institute of Architecture and Technology, in 2003, the master's degree in water conservancy and hydropower engineering from the Hebei University of Engineering, in 2006, and the Ph.D. degree in mechanical engineering from the Beijing University of Technology, in 2015.

From 2006 to 2009, she was a Faculty Member with the School of Mechanical and Equipment Engineering, Hebei University of Engineering. Since 2018, she has been an Associate Professor with the Department of Mechanical and Electrical Engineering, Hebei University of Engineering. She is the author of two research projects, four articles, and one invention patent. Her research interests include hydraulic pneumatic transmission, robotics, and mechanical design theory and methods.

GUANGYU CHEN was born in Hebei, China, in 1995. He received the bachelor's degree in mechanical design, manufacturing and automation from the Hebei University of Engineering, in 2015. He is currently pursuing the master's degree in mechanical engineering with the School of Mechanical and Equipment Engineering, Hebei University of Engineering.

His research interests include machine vision and deep learning.

JIEHUI LIU was born in Hebei, China, in 1968. He received the bachelor's degree in mechanical design, manufacturing and automation from Xi'an Jiaotong University, in 1990, and the master's degree in mechanical design, manufacturing and automation from Tsinghua University, in 2003.

From 1990 to 1998, he was an Engineer with the Institute of Composite Materials and Structures, Hanguang Machinery Factory, a state-owned enterprise of China Shipbuilding Industry Corporation. Since 2006, he has been a Professor and a Graduate Tutor with the Department of Measurement and Control Technology and Instrumentation, School of Mechanical and Equipment Engineering, Hebei University of Engineering. He is the author of six scientific research projects, six articles, and one invention patent. His research interests include mechatronics, mechatronics system control technology, measurement and control technology, and precision mechanical design. He has two scientific and technological awards.

JINXI GUO was born in Henan, China, in 1988. He received the bachelor's degree in mechanical design, manufacturing and automation from the Luoyang Institute of Technology, in 2011, the master's degree in mechanical design and theory from the Hebei University of Engineering, in 2014, and the Ph.D. degree in detection technology and automation devices from the China University of Mining and Technology, in 2018.

Since 2018, he has been a Lecturer and a Graduate Tutor with the School of Mechanical and Equipment Engineering, Hebei University of Engineering. He is the author of one research project and three articles. His research interests include condition automation, filling equipment and control systems, precision instrument design, and virtual instrument system design.