

Study on the surface crack monitoring method of conveyor belt under flutter condition**Bai Wen Luo^a, Xiao Bin Jia^{a*}, Xiao Xin Zeng^b, Xu Dong Li^b and Ting Ting Liao^b**^aCollege of Mechanical and Electrical Engineering, Hunan University of Science and Technology, Xiangtan Hunan, China^bMCC Changtian International Engineering Co., Ltd, Changsha, Hunan, China**ARTICLE INFO***Article history:*

Received 18 December 2024

Accepted 8 March 2025

Available online

8 March 2025

Keywords:

Conveyor belt

Machine vision

Crack detection

Vibration damping device

Spring

YOLOv7

ABSTRACT

In order to monitor the surface condition of the conveyor belt in the process of running, a method of beam structure light irradiation based on machine vision is adopted. A spring-type mechanical vibration damping device is designed to improve the focusing quality of the camera, and an algorithm is proposed to solve the selection of spring parameters under different flutter amplitudes. YOLOv7 deep learning algorithm was adopted and ACmix attention mechanism was introduced to identify the surface cracks of conveyor belt. The experimental results show that the improved YOLOV7-ACmix algorithm can effectively improve the accuracy and generalization ability of image recognition.

© 2025 Growing Science Ltd. All rights reserved.

1. Introduction

The belt conveyor is a crucial piece of equipment for transporting materials over long distances and high volumes, widely used in ports, mining sites, and construction material facilities. Since the conveyor belt is primarily made of rubber, it is susceptible to damage from materials during loading or misalignment, which can lead to severe issues such as tearing or breaking of the belt, resulting in significant economic losses and potential threats to personal safety. Therefore, timely inspection of the belt surface and identification of safety hazards is essential. Traditional manual inspection methods require extensive human resources and may be affected by subjective biases, compromising the quality of inspections. With advancements in industrial intelligent technology, machine vision is now commonly used for online monitoring of belt surface conditions (Fei, 2023).

During the operation of a belt conveyor, factors such as transmission speed, uneven material load, belt elasticity deformation, and mechanical vibrations inevitably cause belt vibrations (YanBin, 2022). These vibrations can significantly impact the focus of industrial cameras during online imaging, making it challenging to obtain stable and clear images of cracks. Therefore, implementing effective vibration reduction measures is crucial for successful machine vision monitoring (Fan, 2023, Qi, 2022). This study examines a material conveyor belt from a pelletizing workshop at Hunan and involves setting up an experimental platform to explore vibration reduction methods, specifically using springs. On this vibration-controlled platform, a linear laser was installed, and an industrial camera was employed to capture clear images of surface cracks on the belt (Haibin et al., 2013), aiding in the development of monitoring algorithms.

In the study of belt surface crack detection algorithms, image recognition algorithms based on artificial intelligence and deep learning are widely used in the field of machine vision for online monitoring (Gu & Zong, 2022). This paper employs the YOLOv7 (Wang et al., 2023) algorithm for object detection and introduces ACmix (Pan et al., 2022) into the network architecture of the object detection algorithm. By combining the advantages of convolution and self-attention mechanisms, the algorithm's sensitivity to small target objects is enhanced. This approach is utilized to achieve intelligent online monitoring of surface cracks on the belt.

* Corresponding author.

E-mail addresses: 2919161158@qq.com (X.B. Jia)

ISSN 2291-8752 (Online) - ISSN 2291-8744 (Print)

© 2025 Growing Science Ltd. All rights reserved.

doi: 10.5267/j.esm.2025.3.002

2. Construction of the Conveyor Belt Testing Platform

The conveyor belt crack online monitoring test platform described in this paper is shown in **Fig. 1**. It includes a conveyor belt system equipped with sensors and actuators to simulate real-world operating conditions. The platform features adjustable components for tuning vibration parameters and integrating different damping technologies. Additionally, it is fitted with a machine vision system to capture and analyze belt surface images, facilitating the assessment of the effectiveness of vibration control methods.

The belt is made of black rubber, with a width of 650 mm and a drive power of 3 kW, utilizing variable frequency speed control with a maximum belt speed of 1.6 m/s. The belt is connected using mechanical chain links, which cause significant vibration every time the belt passes over rollers and tensioning idler. However, this vibration is patterned and can be distinguished from abnormal vibrations.

Two locations for vibration reduction devices are reserved at the center of the conveyor's lower crossbeam. On one side of the vibration reduction device, a camera bracket and a linear laser emitter bracket are installed. The camera used for capturing belt crack images is a high-resolution Hikvision MV-CH250-90UC model (with a resolution of 25 megapixels and a frame rate of 4.5 fps).



Fig. 1. Conveyor belt test platform physical picture

3. Vibration Reduction Device Testing Platform Construction

3.1 Design of Vibration Reduction Device Structure

The structure of a single vibration reduction device is illustrated in **Fig. 2**.

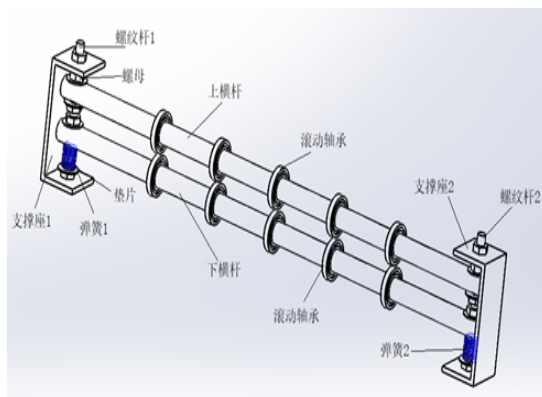


Fig. 2. Structure drawing of vibration damping device



Fig. 3. Vibration reduction device layout diagram

The device consists of an upper and lower crossbar, each equipped with a rolling bearing fixed at regular intervals along their axial direction. The upper crossbar is securely fixed in place by a screw-nut mechanism, while the lower crossbar is connected in series with a spring on the screw and is constrained by upper and lower nuts to limit its range of movement. The conveyor belt passes through the space between the upper and lower crossbars. Under the force of the spring, the rolling bearings on the lower crossbar push the belt into contact with the rolling bearings on the upper crossbar, ensuring that the bearings maintain constant pressure contact with the belt. Two sets of these vibration reduction devices are arranged at a

specific distance, as shown in **Fig. 3**. This distance is optimized for the best image capture range of the industrial camera. The vibration reduction devices are mounted between two lower support beams of the belt conveyor, thereby limiting the vertical vibrations of the belt. The design also takes into account material durability, ease of installation, and compatibility with the conveyor system. Adjustable elements are incorporated to fine-tune the device's performance according to the specific vibration characteristics of the conveyor system. Overall, the goal is to enhance the stability and longevity of the conveyor belt by minimizing the impact of vibrations (Haibin, 2011).

3.2 Static analysis of vibration damping device

When the belt passes between the upper and lower crossbars, the lateral forces generated by its vibrations directly act on both crossbars. The position of the upper crossbar is restricted by the nuts, while the maximum position of the lower crossbar is limited by the nuts, and the minimum position is supported by the pre-tension force of the spring, which acts in the opposite direction to the lateral forces exerted on the lower crossbar by the belt. Since the lower crossbar is in close contact with the belt, the lower crossbar and the belt can be considered as a single unit (Zhu et al., 2018; Wei & LiQun, 2006;). Based on this setup, the static equilibrium equation can be formulated as follows:

$$F_1 + 2Kx + F_3 \cos \theta = F_2 + mg \quad (1)$$

In the equation, F_1 is the preload force of the spring, K is the spring's stiffness coefficient, x is the maximum amplitude, F_3 is the tension force in the belt, $\cos \theta$ is the angle between the belt and the vertical direction when the maximum amplitude is reached, and F_2 is the weight of the lower crossbar. m is the mass per unit length of the belt, and g is the acceleration due to gravity. Due to the varying vibration amplitudes of the conveyor belt at different lengths of the support beam, and since the vibration reduction devices are installed at the same horizontal height along the beam, it is necessary to adjust the height of the upper and lower crossbars using screw-nut mechanisms to accommodate the belt's different vertical positions. Given that the vibration amplitudes and impulses of the belt vary at different positions along the beam, different spring pre-tension forces are required to control the belt's vibration. Therefore, designing suitable springs is a key focus and critical technology in the development of this vibration reduction device.

3.3 Spring parameter calculation

To measure the vibration amplitude and frequency of the belt, the conveyor belt testing platform was used to capture the edge vibrations of the belt at various speeds, with numerical calibration performed. The belt vibration data collected at different speeds is presented in **Table 1**.

Table 1

Vibration parameters of the conveyor belt before the vibration damping device is installed

| Converter frequency (Hz) | Tape speed (mm/s) | Lower belt amplitude Max (mm) | Maximum wave frequency (Hz) |
|-----------------------------|----------------------|-------------------------------------|--------------------------------|
| 15 | 480 | 15 | 4 |
| 20 | 640 | 16 | 10 |
| 25 | 800 | 21 | 10 |
| 30 | 960 | 15 | 13 |
| 35 | 1120 | 10 | 10 |
| 40 | 1280 | 30 | 10 |
| 45 | 1440 | 9 | 7 |

Based on the static equilibrium equation (Eq. (1)) and using the principle of energy conservation, the following differential equation can be established:

$$\frac{1}{2} \rho \dot{x}^2 = E_p + E_D + \frac{1}{2} Kx^2 \quad (2)$$

In the equation, E_p represents the gravitational potential energy added to the system when the belt's vibration amplitude is at its maximum. E_D denotes the work done by the elastic force generated by the belt deformation, and K is the spring's stiffness coefficient. The terms E_D and E_p are neglected

$$\frac{1}{2} \rho \dot{x}^2 = \frac{1}{2} Kx^2 \quad (3)$$

Solving this equation yields:

$$K = \rho \ln^2 x \quad (4)$$

Using this calculation method, the spring stiffness coefficient is determined as shown in **Table 2**.

Table 2
Calculated values of spring stiffness coefficient

| Converter frequency (Hz) | 15 | 20 | 25 | 30 | 35 | 40 |
|--------------------------|------|------|------|------|------|------|
| K (N/mm) | 26.4 | 25.6 | 22.3 | 26.4 | 31.6 | 22.9 |

The required spring stiffness coefficient is 31.6 N/mm. Based on the parameters mentioned above, the maximum lateral load exerted by the belt is approximately 350 N. Finally, referring to the Mechanical Design Handbook and considering practical conditions, the specific spring parameters are listed in **Table 3**.

Table 3
Spring Types and Dimensional Parameters

| Spring type | mean diameter of coil | Material diameter | pitch | number of active coils | number of total coils | unsupported height |
|----------------------|-----------------------|-------------------|--------|------------------------|-----------------------|--------------------|
| Top tight and smooth | 20 mm | 3 mm | 6.4 mm | 6.5 | 8.5 | 46 mm |

4. Belt Crack Machine Vision Monitoring Methods

As shown in **Fig. 4**, the machine vision system described in this paper consists of an industrial camera and a linear laser emitter, both mounted on the vibration reduction device platform. The linear laser emitter projects a bright green "I"-shaped stripe onto the belt surface. When a crack on the belt surface passes through the laser stripe, it causes a noticeable "interruption" in the laser stripe image, as illustrated in **Fig. 5**. However, variations in brightness or material adhesion on the belt surface can also cause similar "interruptions" in the laser stripe image. Therefore, how to quickly identify the real crack of the belt is also an important research content of this study. This paper uses the YOLOv7 deep learning algorithm for crack detection on the conveyor belt.

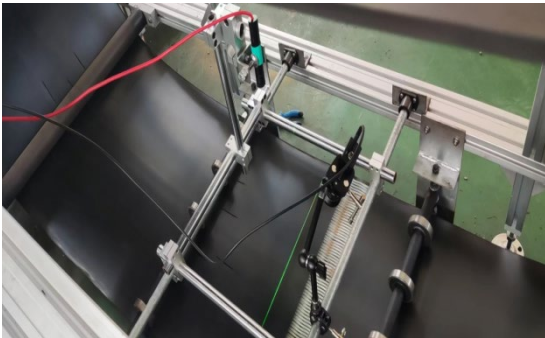


Fig. 4. Machine vision device layout diagram

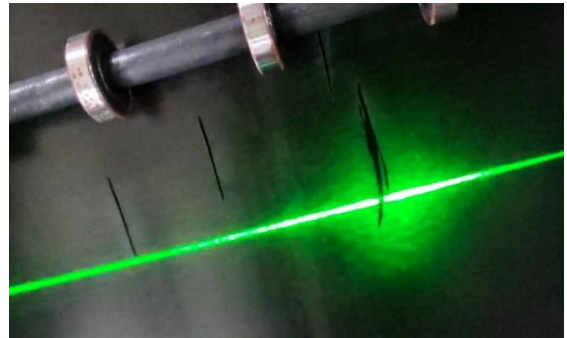


Fig. 5. Linear green laser projection

YOLOv7 is one of the most advanced single-stage object detection algorithms, capable of meeting both real-time and high-accuracy requirements for belt surface crack detection. It is an enhancement of YOLOv5, featuring several improvements. The primary framework of YOLOv7 consists of three parts: the input, the backbone network, and the head. Compared to YOLOv5, YOLOv7 incorporates the E-ELAN network module, which enhances training and inference efficiency. Additionally, the introduction of auxiliary heads improves detection accuracy without affecting inference time. However, due to the elastic nature of the belt, cracks may contract, leading to smaller detection targets and issues such as reduced detection accuracy and missed small targets (Chen et al., 2017; Gao et al., 2021). To address these challenges, the YOLOv7 network is improved by incorporating the ACmix attention mechanism. Convolution (Krizhevsky et al., 2017; LeCun et al., 1989) and self-attention (Vaswani, 2017) are two powerful modules used in representation learning. Convolutional modules typically focus on information within a local receptive field, whereas self-attention modules consider information across the entire image (Niu et al., 2021). This makes self-attention mechanisms more flexible compared to convolutional neural networks (CNNs). However, this flexibility comes with a trade-off: self-attention mechanisms often require more training data to perform effectively (Cordonnier et al., 2019). In scenarios with limited training data, self-attention mechanisms may suffer from overfitting.

On the other hand, CNNs generally require less training data and are less prone to overfitting. However, they may not fully leverage the benefits of large datasets as effectively as self-attention mechanisms can. To leverage the strengths of both types of neural networks, this paper introduces the ACmix attention mechanism. The ACmix module elegantly combines convolutional and self-attention mechanisms, harnessing the benefits of both. The principle is as follows:

ACmix merges the similar first stages of CNNs and self-attention networks. In the second stage, it creates two branches: one for convolutional processing and another for self-attention processing. Each branch performs its respective computations in the second stage. Finally, the outputs from both branches are combined through weighted fusion, as illustrated in **Fig. 6**.

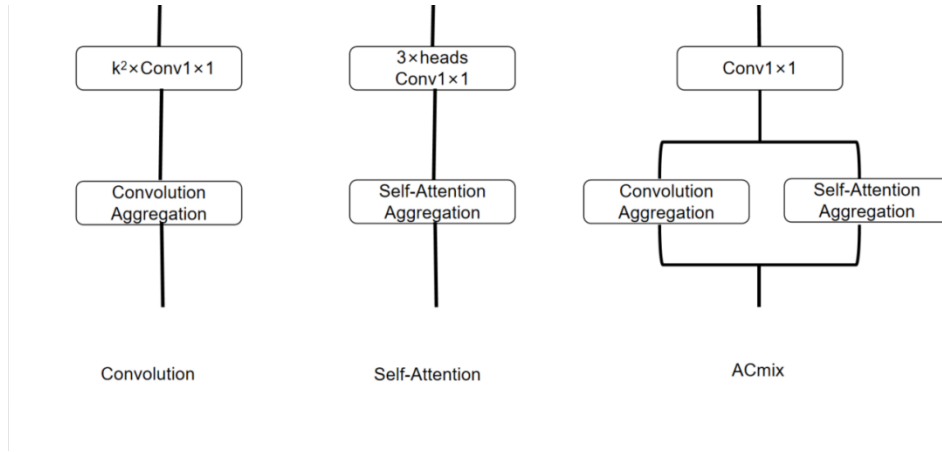


Fig. 6. Three kinds of network structure diagram

After merging the outputs from the two branches, the ACmix attention mechanism integrates the strengths of both convolutional and self-attention networks. This integration enhances the detection of both global and local features, improving the network's overall object detection performance.

We abstract the convolution process into two stages:

First Stage: Convolution is performed by applying a specific pixel from the convolutional kernel to the previous feature map, generating a feature map unique to that kernel. **Second Stage:** The feature maps obtained in the first stage are shifted according to their respective pixel positions and then summed together, as illustrated in **Fig. 7**.

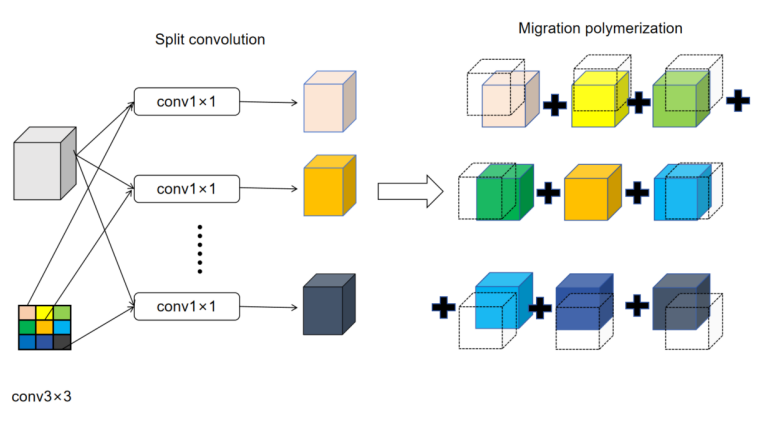


Fig. 7. Convolution schematic

In the first step of the convolution operation, the $k \times k$ convolution kernel is decomposed into $K \times K$ 1×1 convolution kernels. Each element of these 1×1 kernels is multiplied by the corresponding pixel value from the previous layer's feature map, without performing any summation. The operation can be described as follows:

$$\tilde{g}_{ij}^{(p,q)} = K_{p,q} f_{ij} \quad (5)$$

In the second step, the feature maps obtained from the first step are shifted according to their corresponding positions. This shifting process can be represented by the following formula:

$$g_{ij}^{(p,q)} = Shift(\tilde{g}_{ij}^{(p,q)}, p - \frac{k}{2}, q - \frac{k}{2}) \quad (6)$$

After obtaining the shifted feature maps, the next step is to aggregate these features. The aggregation process is given by the following formula:

$$g_{ij} = \sum_{p,q} g_{ij}^{(p,q)} \quad (7)$$

The attention mechanism first calculates the weights for the feature at the current pixel (i, j) relative to the features in its surrounding area. It then performs a weighted sum of these features. This process can be visualized as shown in Fig. 8.

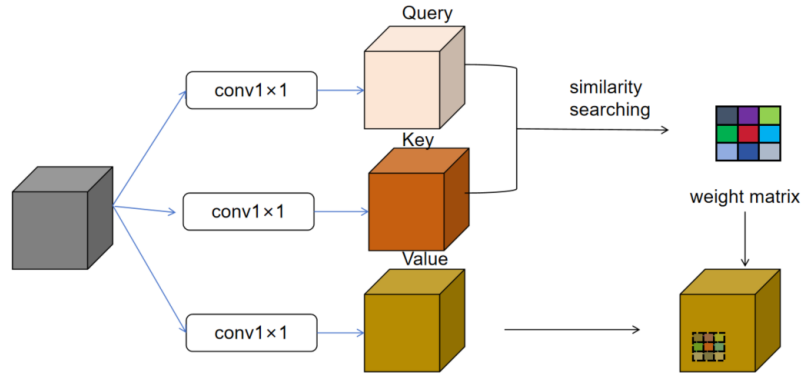


Fig. 8. Self-attention mechanism schematic

In the first step of the attention mechanism, the input features are linearly transformed into three different representations: Query (Q), Key (K), and Value (V). This transformation can be expressed by the following formulas:

$$q_{ij}^{(l)} = W_q^{(l)} f_{ij}, k_{ij}^{(l)} = W_k^{(l)} f_{ij}, v_{ij}^{(l)} = W_v^{(l)} f_{ij} \quad (8)$$

The F_{ij} is the input feature matrix. In the second step of the attention mechanism, the attention weights are calculated and then used to aggregate the Value (V) matrix, effectively focusing on relevant features. This process involves the following steps and formulas:

$$g_{ij} = \left\| \left(\sum_{l=1}^N \sum_{a,b \in N_k(i,j)} A(q_{ij}^{(l)}, k_{ab}^{(l)}) v_{ab}^{(l)} \right) \right\| \quad (9)$$

$N_k(i, j)$ denotes a pixel region centered at (i, j) with a spatial width of k . $A(W, k)$ represents the weights corresponding to the $N_k(i, j)$ region.

5. Experimental results and analysis

5.1 Vibration reduction effect experiment

Based on the vibration reduction device structural design and the spring types and dimensions detailed in Table 3, the vibration reduction devices were installed at 600 mm intervals along the center of the lower belt on the conveyor test platform for laboratory experiments. Using the same belt speed and the same edge vibration measurement methods and equipment as described in Table 1, the maximum amplitude data of the belt was obtained, as shown in Table 4.

Table 4

Amplitude of conveyor belt after vibration damping device installation

| Converter frequency (Hz) | Tape speed(mm/s) | Lower belt amplitude Max(mm) | damping range (%) |
|-----------------------------|------------------|------------------------------|-------------------|
| 15 | 480 | 4 | 73 |
| 20 | 640 | 5 | 69 |
| 25 | 800 | 6 | 71 |
| 30 | 960 | 5 | 67 |
| 35 | 1120 | 5 | 50 |
| 40 | 1280 | 6 | 80 |

From Table 4, it can be observed that after installing the vibration reduction devices, the average maximum amplitude of the lower belt was 5 mm, with an average reduction in vibration amplitude of 66.7%. The factors affecting this result include uneven belt tension distribution and collisions between the belt chain links. However, since the longitudinal tear cracks in the belt have a simple profile, even with a 10 mm focusing error in the camera lens, relatively clear images of the belt cracks can still be obtained.

5.2 yolov7 improved algorithm detection effect experiment

A total of 1,036 images were collected in this study. The images were cropped to a 1:1 ratio, resulting in a resolution of 480×480 pixels, and then augmented to 2,038 images. Among these, 2,004 images were used for training, and 154 images

were used for validation. The platform utilized for operations included a 13th Gen Core i5-13490 CPU, NVIDIA GeForce RTX 3060 Ti GPU, with programming conducted in Python 3.11.3, running on Windows 11. The environment also included Anaconda3 and the PyTorch deep learning framework, with CUDA used for acceleration.

To validate the advancement of the YOLOv7 algorithm enhanced with the hybrid attention mechanism, it was compared with other YOLO series algorithms (YOLOv7, YOLOv7-SE (Hu et al., 2018), and YOLOv5s) on the same dataset. Training was conducted for 150 iterations, and the best-performing model for each algorithm was saved. The evaluation metrics used include precision, recall, F1 score, mAP@0.5, and detection speed. mAP@0.5 represents the mean average precision when the Intersection over Union (IoU) threshold is set at 0.5, while the F1 score is a measure that combines precision and recall into a harmonic mean. The model parameters are listed in the table below:

Table 5

Comparison of algorithm models

| | Precision | Recall | F1 | map@0.5 |
|--------------|-----------|--------|-------|---------|
| Yolov7-acmix | 88% | 93.02% | 0.904 | 83.7% |
| Yolov7 | 83% | 85.06% | 0.839 | 81.3% |
| Yolov7-se | 81% | 95.4% | 0.874 | 75.3% |
| Yolov5s | 80% | 80.49% | 0.802 | 76.5% |

Compared to other methods, the YOLOv7 model with the integrated hybrid attention mechanism shows significant performance improvements. Specifically: Precision increased by 5 percentage points compared to the original YOLOv7. Recall improved by 7.96 percentage points compared to YOLOv7. mAP@0.5 rose by 2.4 percentage points compared to YOLOv7. When compared to YOLOv5s: Precision increased by 8 percentage points. Recall improved by 12.4 percentage points. mAP@0.5 increased by 7.2 percentage points. Compared to YOLOv7-SE, which includes the SE attention mechanism: Precision and mAP@0.5 both showed improvements. Recall saw a slight decrease. Overall, the YOLOv7-ACMix model outperforms mainstream algorithms across all detection metrics. Additionally, it maintains an average detection speed of 60 fps, meeting the requirements for real-time monitoring.

To further validate the improved model's detection performance, images from the test set under two different lighting conditions—well-lit and dimly lit scenes—were selected for detection experiments. The results are shown in **Fig. 9** and **Fig. 10**. The figures show the predictions of longitudinal tears on the belt surface by the improved algorithm and other models after training.

In **Fig. 9**, the clear breaks in the laser pattern are visible in the well-lit scene, and all four algorithms can detect the laser breaks. However, the proposed algorithm shows slightly higher confidence compared to the others. In **Fig. 10**, under dim lighting, the belt tears can only be identified through the laser breaks. The proposed algorithm is the only one that accurately detects these breaks under such conditions. The other three algorithms missed detections, with the confidence of the detected break on the far right being much higher for the proposed algorithm. Overall, YOLOv7-ACMix accurately locates and identifies laser breaks in both well-lit and dimly lit scenarios, without any missed or false detections. The experimental results demonstrate that the improved model effectively adapts to different detection scenarios and exhibits strong generalization capabilities.



Fig. 9. the detection results from left to right are as follows: YOLOv7-ACMix, YOLOv5s, YOLOv7-SE, and YOLOv7.

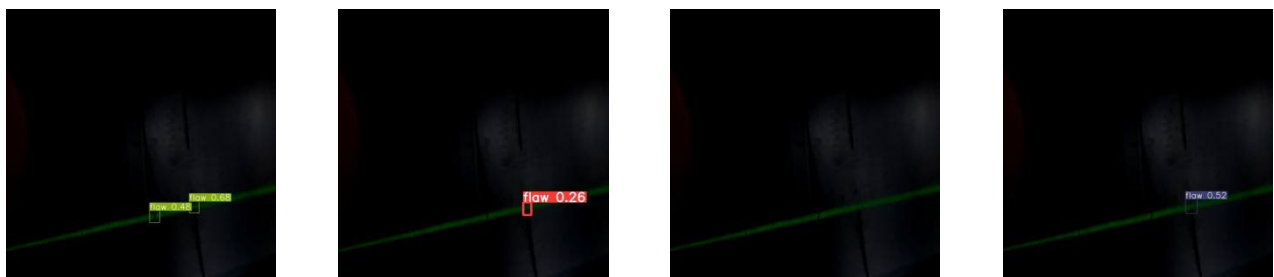


Fig. 10. the detection results from left to right are as follows: YOLOv7-ACMix, YOLOv5s, YOLOv7-SE, and YOLOv7.

6. Conclusions

(1) The machine vision method is used to monitor the surface condition of the conveyor belt online, and the test platform is set up according to the actual operation parameters. The test platform adopts a single camera and a laser irradiation mode of a frame structure.

(2) In view of the difficulty of camera focusing due to flutter during the operation of the conveyor belt, a set of spring vibration damping devices was designed and a test platform was built, through which the damping effectiveness and parameter selection calculation of the spring were studied.

(3) YOLOv7 deep learning detection algorithm and ACmix attention mechanism are adopted to improve the accuracy and generalization ability of image recognition in the field of surface crack identification of conveyor belt. Thousands of images are collected on the test platform for sample training. The experimental results verify that the improved YOLOv7-ACmix algorithm has improved compared with the mainstream algorithm in various detection standards, and the average detection speed can meet the technical indicators of real-time monitoring.

References

- Chen, C., Liu, M. Y., Tuzel, O., & Xiao, J. (2017). R-CNN for small object detection. In *Computer Vision—ACCV 2016: 13th Asian Conference on Computer Vision, Taipei, Taiwan, November 20-24, 2016, Revised Selected Papers, Part V 13* (pp. 214-230). Springer International Publishing.
- Cordonnier, J. B., Loukas, A., & Jaggi, M. (2019). On the relationship between self-attention and convolutional layers. arXiv preprint arXiv:1911.03584.
- Fan, L. (2023). On-line monitoring system of belt conveyor operation status. *Petrochemical Technology*, 30(05), 238-239.
- Fei, W. (2023). Analysis and Treatment of Common Faults in Coal Mine Belt Conveyors. *Mechanical Management and Development*, 38(2), 245-246+251.
- Gao, X. B., Mo, M., Wang, H., & Leng, J. (2021). Recent advances in small object detection. *Journal of Data Acquisition and Processing*, 36(3), 391-417.
- Gu, Y. L., & Zong, X. X. (2022). A review of object detection study based on deep learning. *Modern Information Technology*, 6(11), 76-81.
- Haibin, L., Chunming, Z., & Yuanzheng, Z. (2011). A new inspection method for rip of conveyor belt based on line laser stripe. *Optical Technique*, 37(4), 466-470.
- Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 7132-7141).
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6), 84-90.
- LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., & Jackel, L. D. (1989). Backpropagation applied to handwritten zip code recognition. *Neural computation*, 1(4), 541-551.
- Niu, Z., Zhong, G., & Yu, H. (2021). A review on the attention mechanism of deep learning. *Neurocomputing*, 452, 48-62.
- Pan, X., Ge, C., Lu, R., Song, S., Chen, G., Huang, Z., & Huang, G. (2022). On the integration of self-attention and convolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 815-825).
- Qi, Wu. (2022). Operation Monitoring Technology of Mining Belt Conveyor Rollers. *Mechanical Management and Development*, 37(01), 133-142.
- Vaswani, A. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*(pp.5998-6008).
- Wang, C. Y., Bochkovskiy, A., & Liao, H. Y. M. (2023). YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 7464-7475).
- Wei, Z., & LiQun, C. (2006). Transverse vibration control of an axially moving string system: Energy method. *Journal of Mechanical Strength*, 28, 201-204.
- YanBin, P. (2022). Fault analysis and treatment measures of belt conveyor. *Metallurgy and Materials*, 42(04), 96-97.
- Zhu, Y., Zheng, C.L., & Liu, Z. (2018). Analysis of transverse vibrations of moving belt caused by fulcrum motion. *Noise and Vibration Control*, 38(01). 90-170.

