China Instrument and Control Society
Instrumentation
INSTRUMENTATION, Vol. 13, No. 2, June 2026
https://doi.org/10.15878/j.instr.202600338

Article

Real-Time Detection and Assessment of Longitudinal Tears in Mining Conveyor Belts Using Line Lasers and YOLO-DM

Hongtao Hao[1*], Mai Xuewu[1], Chenhe Liu[1], Pengfei Teng[1], Xiaodong Ma[1], Baozhong Feng[2]

[1] School of Mechanical Engineering, Ningxia University, Yinchuan 750021, China

[2] Ningxia Tiandi Northwest Coal Machinery Co., Ltd., Shizuishan 753000, China

[*] Corresponding author email: haoht_03@126.com

Abstract: Conveyor belts serve as critical components in bulk material transportation systems, wherein longitudinal tearing constitutes one of the most severe failure modes jeopardizing operational safety. However, current detection approaches commonly face limitations in incipient tear recognition, real-time deployment, and rigorous quantitative evaluation. This study develops an integrated framework that couples single-line laser projection with an enhanced YOLOv11n-based detector (YOLO-DM) and an online tear-length estimation algorithm. YOLO-DM introduces two key advancements: (1) the Dual Enhanced Squeeze-and-Excitation (DESE) attention module embedded in the backbone, in which SE (capturing global channel dependencies) and ECA (modeling local channel interactions) branches operate in parallel and are adaptively fused through learned weights to reinforce subtle tear cues; (2) a Multi-Scale Laser Line Feature Fusion (MSLF) module incorporated into the neck, featuring parallel 1×1/3×3/5×5 convolutions together with a direction-aware asymmetric convolution branch to highlight laser-stripe discontinuities, followed by scale-adaptive weighted fusion. For quantitative assessment, a spatiotemporal tracking-based duration model with boundary compensation is combined with belt-speed integration to enable online tear-length estimation. Experiments on a 3,000-image dataset show that YOLO-DM achieves 98.7% mAP@0.5 at 58.1 FPS, while tear-length estimation attains 95.8% accuracy, satisfying the requirements for real-time monitoring in mining conveyor-belt applications.

Keywords: Conveyor belts; detection of longitudinal tears; line laser technology; YOLO-DM; Measurement of tear length

Citation: Hongtao Hao, Mai Xuewu, Chenhe Liu, Pengfei Teng, Xiaodong Ma, Baozhong Feng."Real-Time Detection and Assessment of Longitudinal Tears in Mining Conveyor Belts Using Line Lasers and YOLO-DM." Instrumentation 13, no.2 (June 2026). https://doi.org/10.15878/j.instr.202600338

1 Introduction

Belt conveyors are indispensable in coal-mine transportation systems due to their high throughput, long conveying distances, and stable continuous operation[1,2]. Nevertheless, belts operating under harsh underground conditions are vulnerable to diverse forms of damage, among which longitudinal tearing is one of the most severe failure modes[3,4]. Such tearing typically occurs when sharp foreign objects puncture or abrade the belt surface, initiating cracks that extend along the longitudinal direction[5]. If not identified in time, a small defect can propagate rapidly, resulting in catastrophic belt failure, substantial economic losses, and serious safety risks[6,7]. Therefore, an accurate, reliable, and real-time longitudinal-tear detection system is essential for safe and efficient coal-mine production.

Current longitudinal-tear detection approa- ches can be broadly grouped into contact- based methods and non-contact methods.

1.1 Contact-Based and Traditional Vision Methods

Early tear detection relied mainly on contact- based devices, including rod-type detectors, wire- type detectors, and swing-roller detectors[8]. Although structurally simple, these solutions typically exhibit limited detection accuracy, pronounced mechanical wear, and reduced reliability under coal-dust accumulation [9]. As a result, they have gradually been superseded by non-contact technologies.

Traditional vision-based methods directly analyze visible-light or infrared images of the belt surface. Peng et al. [10] proposed image segmenta- tion and edge-detection schemes; however, their performance is highly sensitive to illumination changes. Qiao et al.[11] developed infrared imaging systems with improved robustness, albeit at the expense of costly thermal equipment. In general, these approaches depend on fixed thresholds or handcrafted features, providing limited adaptabi- lity to variable underground environments and often producing elevated false-alarm rates[12].

1.2 Structured Light-Based Detection Methods

To mitigate the limitations of passive imaging, structured-light (laser-line projection) techniques have been introduced to actively illuminate the belt surface, converting surface deformation into geometric distortion of projected laser stripes[13,14]. Li et al. [15] developed a single-laser-line system using the Steger algorithm for centerline extraction, demonstrating feasibility but offering limited coverage and suffering missed detections at high belt speeds. Xu et al.[16]proposed a multi-line laser method with three parallel stripes and grouped center-point extraction to achieve tear tracking, while its generalization remained constrained by environmental sensitivity. Lv et al. [17]introduced an improved gray-centroid approach for centerline extraction, but stripe breakage was observed at large trough depths. Zhao et al. [18] designed a LoG-preprocessing- based algorithm, reporting 95.14% precision and 92.63% recall in industrial deployment.

Despite these advances, structured-light methods still face persistent challenges, including inadequate robustness to environmental interfer- ence, difficulty in capturing minor incipient tears, and limited real-time capability on embedded platforms[19].

1.3 Deep Learning and Integrated Approaches

Recent progress in deep learning has substant- ially advanced industrial defect detection [20,21]. Wang et al.[22] employed CNNs to classify belt- damage types, showing strong representation learning but requiring large labeled datasets. Zhang et al.[23] developed an improved YOLOv3 framework with feature pyramid networks, though inference was relatively slow. Yu et al. [24] presented a dimension-reduced YOLOv5 variant (R-YOLOv5) tailored for line-structured-light imagery, achieving 95.83% accuracy but struggling to detect tears narrower than 12 pixels.

To further enhance performance, increasing attention has been given to integrating structured light with deep learning[25,26]. Che et al. [27] proposed an audio – visual fusion method with 96.23% accuracy, but it increased hardware complexity. Li et al. [28] introduced a multi-set laser-line-assisted detection strategy with 95.83% accuracy, yet it required manual threshold tuning. Zhou et al. [29] applied CBAM – YOLOv5 for multi- class damage detection, whereas performance degraded under unfavorable illumination.

Overall, existing methods still encounter critical limitations:

(1) Limited detection coverage: insufficient spatial sampling can cause missed detections across the belt width.

(2) Weak small-target sensitivity: early-stage tears are frequently overlooked due to inadequate feature representation.

(3) Unsatisfactory real-time performance: high computational cost restricts deployment on embedded platforms.

(4) Lack of quantitative evaluation: many systems only report tear presence without precise measurements.

(5) Poor environmental adaptability: coal dust and uneven illumination degrade accuracy.

1.4 Research Contributions

To address these issues, this studyproposes a real-time longitudinal-tear detection and evaluation system that integrates single-line laser projection with an enhanced YOLOv11n-based model (YOLO-DM). The main contributions are as follows:

(1) Specialized architecture for laser-stripe features: Many existing studies directly apply generic object detectors to laser-line images without explicitly modeling the linear geometric nature of laser stripes and the abrupt pattern changes induced by tearing. We design a DESE dual-attention mechanism and an MSLF multi-scale laser-line feature fusion module, tailored to the position shifts, intensity variations, and morphological distortions of laser stripes, enabling efficient feature extraction and accurate recognition[30].

(2) Scale-adaptive fusion for multi-scale laser-stripe cues: Conventional structured-light methods often depend on fixed thresholds or manually tuned parameters, which limits adaptability across tear scales and environmental conditions. The proposed MSLF module incorporates direction-aware enhancement and scale-adaptive selection to adjust fusion weights dynamically according to tear geometry, thereby improving sensitivity to minute tears.

(3) Integrated tear-length quantification: Most existing approaches focus on tear detection without precise length estimation. We propose a tear-length

calculation algorithm based on spatiotemporal target tracking and belt-speed integration, providing accurate quantitative results to support maintenance decision-making.

1.5 Paper Organization

The remainder of this paper is organized as follows. Section 2 reveals the intrinsic pattern of laser-stripe variations by establishing a geometric discontinuity model for line lasers and analyzing the physical mechanisms of tearing. Section 3 introduces the YOLO-DM detection model, including the enhanced network integrating the DESE dual-attention mechanism and multi-scale feature fusion, as well as the construction of relevant datasets. Section 4 presents a precise tear-length measurement algorithm based on spatiotemporal object association and motion compensation. Section 5 reports experimental results and platform validation. Finally, Section 6 summarizes the paper and outlines future research directions.

2 System Architecture

Designing a longitudinal-tear detection system for conveyor belts requires accommodat- ing the harsh and variable conditions encountered in mining environments, while simultaneously ensuring high detection accuracy and real-time responsiveness. This section presents the key design considerations of the proposed system, which integrates line-laser projection with deep- learning-based perception. In addition, the physical origin of tear-induced laser-stripe deformation is discussed, and the overall system architecture is described.

2.1 Analysis of Detection Principle

(1) Geometric Model of Line-Laser Projection

The proposed system adopts structured-light principles, where a line laser generates a continuous stripe on the conveyor-belt surface to facilitate damage sensing. As shown in Fig. 1, a triangulation setup is formed by the laser emitter, the belt surface, and an industrial camera. The laser projects a line onto the belt at a predefined incidence angle, producing a smooth and uniform stripe under intact conditions. When a longitudinal tear is present, local surface geometry changes disrupt the stripe, leading to measurable deviations in its shape and position. These stripe variations are captured by the camera and converted into digital image signals for subsequent processing.

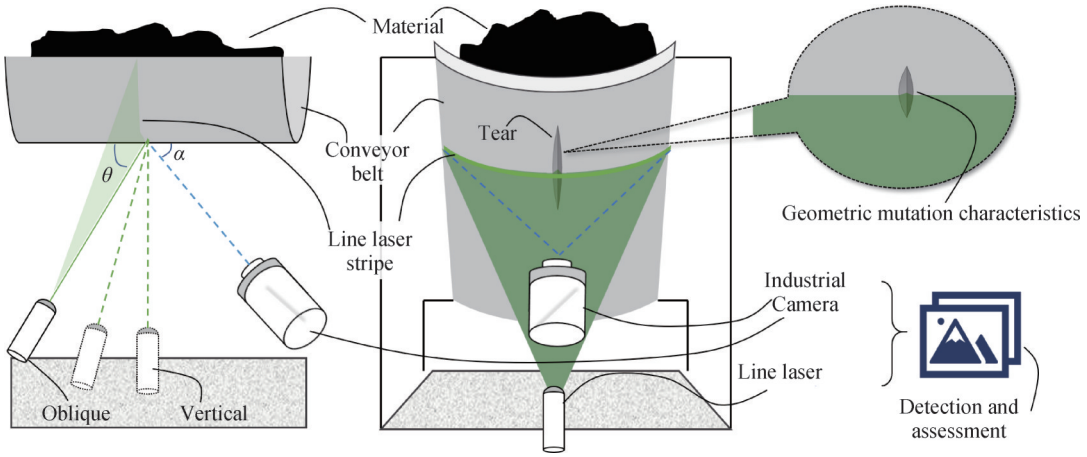

Fig.1 Basic experimental setup diagram

To describe the geometric relationships of the laser projection, a coordinate system is defined. The laser-beam plane is expressed as:

$$\Pi_l: A_l x + B_l y + C_l z + D_l = 0 \tag{1}$$

Under intact conditions, the conveyor-belt surface is approximated as a plane:

$$\Pi_B: A_B x + B_B y + C_B z + D_B = 0 \tag{2}$$

Accordingly, the projected laser stripe on the belt corresponds to the intersection of the two planes, yielding the projection line

$$\mathbf{L}(t) = \mathbf{P}_0 + t\mathbf{d} \tag{3}$$

Where $P0$ denotes a reference point on the laser stripe and d is the stripe direction vector, obtained from the cross product of the two plane normal vectors:

$$\mathbf{d} = \mathbf{n}_L \times \mathbf{n}_B \tag{4}$$

Where $\mathbf{n}_L = (A_L, B_L, C_L)^T$ and $\mathbf{n}_B = (A_B, B_B, C_B)^T$ are the normal vectors of the laser plane and the belt-surface plane, respectively.

The projection angle is a key factor that governs the measurement sensitivity. The feasible angle range is determined via theoretical analysis. As illustrated in Fig. 1, let θ denote the angle between the laser plane and the belt surface (projection angle), and let α denote the angle between the camera optical axis and the belt surface (observation angle). Based on triangulation, the lateral displacement Δx of the laser stripe induced by a tear depth h can be written as:

$$\Delta x = h \cdot (\tan\theta + \tan\alpha) \tag{5}$$

Equation (5) indicates that increasing θ and/or α enlarges the stripe displacement for a given tear depth, thereby improving sensitivity. In practice, however, excessively large angles may (i) cause deep tears to occlude the laser stripe, leading to missed detections; (ii) broaden the projected stripe under oblique incidence, reducing spatial resolution; and (iii) shift the stripe near the belt edges beyond the camera field of view. Therefore, a practical configuration is to employ near-normal laser projection with oblique camera observation. In this work, the recommended settings are a projection angle of 90° ± 15°and an observation angle of 45° ± 5°.

The above geometric formulation provides the theoretical basis for subsequent image processing and feature extraction, enabling the enhanced YOLOv11n model to capture tear-induced stripe deviations more reliably. Moreover, it supplies the geometric constraints required for online tear-length estimation.

(2) Physical Mechanism of Laser Line Shifts Induced by Tearing

The defining characteristic of longitudinal tears in conveyor belts is the development of elongated cracks that propagate along the length of the belt (in the direction of material transport), with depths typically ranging from a few millimeters to several centimeters. As shown in Fig. 2, the tear initiation process can be divided into three distinct phases: (i) microcrack formation due to localized stress concentration, (ii) crack propagation under tensile stress, and (iii) the eventual development of a macroscopic tear. During this progression, the initially flat conveyor-belt surface experiences abrupt changes in height and edge inclination within the tear zone, leading to alterations in both the laser's projection path and its reflection properties.

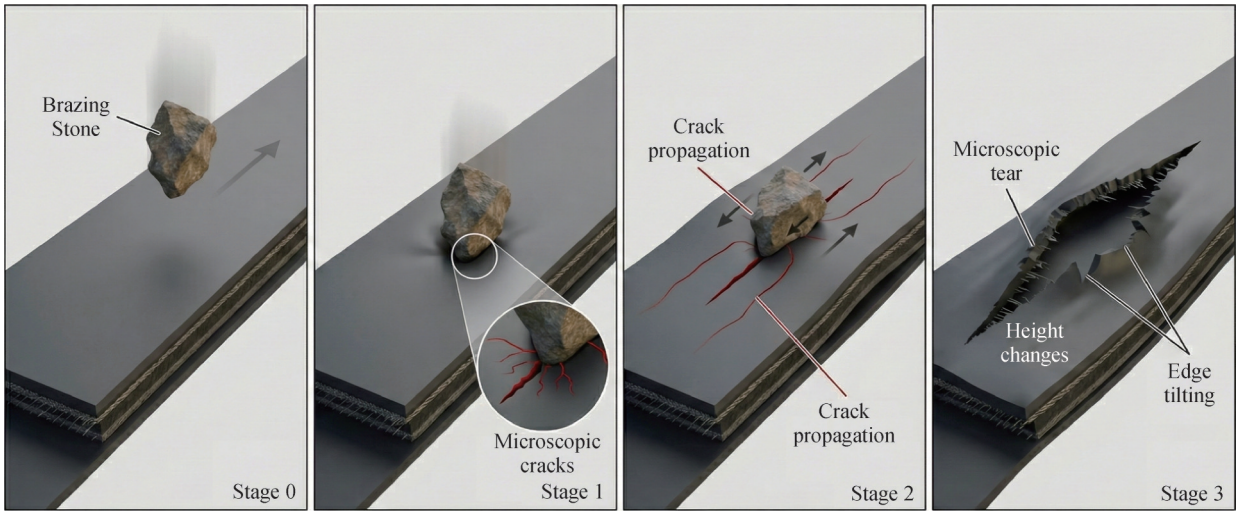

Fig.2 Mechanism of Longitudinal Tearing in Conveyor Belts: Stage 0: Brazing stone fallingonto the conveyor belt; Stage 1: Initial contact andmicro-crack initiation due to stress concentration; Stage 2: Crack propagationunder tensile stress; Stage 3: Final macroscopictear formation with surfaceheight changes

As shown in Fig. 3, tears primarily affect laser stripes in three key ways. The most noticeable effect is positional shift: when the laser beam encounters the depression caused by a tear, the abrupt decrease in surface elevation leads to a clear lateral displacement of the laser line in the image. Intensity variation reflects the interaction between the laser and the irregular surface. The rough edges of the tear scatter the laser light, causing the line to broaden or darken at the tear location. Shadow effects may occur in deep tears, completely interrupting the laser line in certain regions.

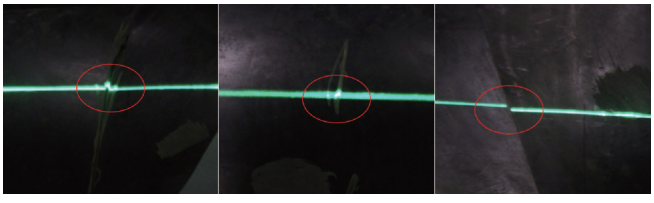

Fig.3 The influence of tearing on laser stripes

These physical phenomena provide valuable feature information for deep learning-based detection algorithms. The YOLO-DM model, with its DESE dual-attention mechanism and MSLF multi-scale laser line feature fusion, effectively captures the positional shifts, intensity variations, and morphological distortions of the laser lines, enabling accurate identification of tears of varying severity. Furthermore, the geometric disruption of the laser lines serves as a reliable spatial reference for the accurate calculation of tear length.

2.2 System Composition and Design

Building on the detection principles outlined earlier, a comprehensive detection system has been designed, integrating line-laser projection, image acquisition, intelligent processing, and result output. Following a modular design approach, each component operates both independently and collaboratively, ensuring the system's reliability and maintainability.

The line laser serves as the core sensing element,

responsible for projecting the laser stripe onto the conveyor belt surface. Given the specific conditions of mining environments, as shown in Fig. 4(a), a 532 nm green laser with a 1 W power output was chosen. The laser projector is mounted at a 45-degree projection angle.

The industrial camera converts the laser stripe data into digital images, with its performance directly affecting both detection accuracy and real-time capability. As depicted in Fig. 4(b), the system employs a high-speed area-scan CMOS camera with a resolution of 1280 × 1024 pixels and a frame rate of 90 fps, fulfilling the image acquisition demands during high-speed conveyor operation.

As illustrated in Fig. 4(c), the conveyor and experimental support frame provide stable mechanical support and a standardized testing environment for the detection system. The test rig uses a steel structure and aluminum-profile framework, ensuring excellent rigidity and vibration resistance. It is capable of supporting equipment such as lasers and cameras. To simulate real mining conditions, the rig is equipped with a high-frequency vibration generator and an adjustable lighting system, enabling the replication of various operational conditions in the laboratory.

The perpendicularity of the laser line relative to the conveyor belt's direction of movement is maintained by securely mounting the laser to rigid aluminum profiles. Based on the projection angle analysis in Section 2.1, the laser has an adjustable swing range of ±15° . During detection, the direction-sensing enhancement mechanism within the MSLF module (described in Section 3.2) continuously monitors the laser's orientation. If any directional deviation is detected, it automatically adjusts the feature extraction weights to maintain detection stability.

The trainable PC-based processing unit acts as the system's intelligent core, handling image processing, model training, inference calculations, and result output. As shown in Fig. 4(d), the processing unit features a high-performance industrial computer with an Intel i7-12700K processor, NVIDIA RTX 4080 GPU, 32 GB of memory, and a 2 TB solid-state drive. This configuration meets the computational demands of deep learning model training and real-time inference.

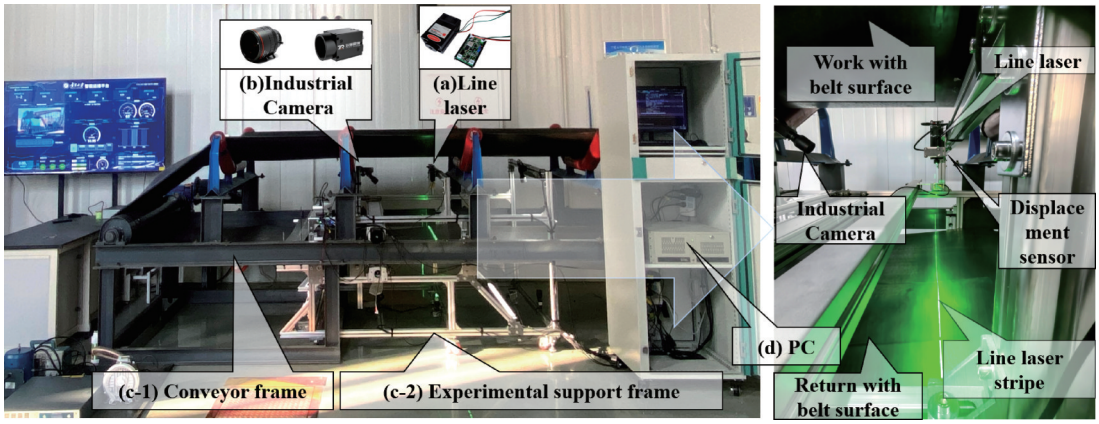

Fig.4 System hardware composition

As shown in Fig. 5, the system workflow illustrates the seamless integration of hardware and software. The laser beam generated by the laser source is modulated by an optical system before being projected onto the conveyor belt surface, creating a light stripe with distinct geometric properties. When tears are present on the belt surface, the shape of the light stripe undergoes corresponding changes. An industrial camera captures real-time images of the laser stripe and transmits the data to the processing unit via a high-speed data interface. The processing unit performs initial image pre-processing, including noise filtering and contrast enhancement, before applying the YOLO-DM model for tear detection. The detection results are then processed by a target tracking algorithm. By integrating conveyor belt speed data, the tear length is calculated. Finally, the detection results and alarm information are displayed on the human-machine interface.

3 YOLO-DM Detection Model

In this study, YOLOv11n was selected as the baseline model for three main reasons.

First, it offers excellent real-time performance. By integrating the improved C3k2 module and the SPPF architecture, YOLOv11n significantly reduces computational complexity while maintaining detection accuracy compared with YOLOv8 and YOLOv10. This characteristic makes it particularly suitable for deployment on embedded platforms.

Second, YOLOv11n demonstrates strong capability in small object detection. Its multi-scale detection heads and enhanced anchor-free mechanism improve the detection of small and fragmented objects.

Third, YOLOv11n has a highly scalable architecture. Its modular design facilitates the integration of the proposed DESE and MSLF modules, enabling customized optimization for laser line feature extraction.

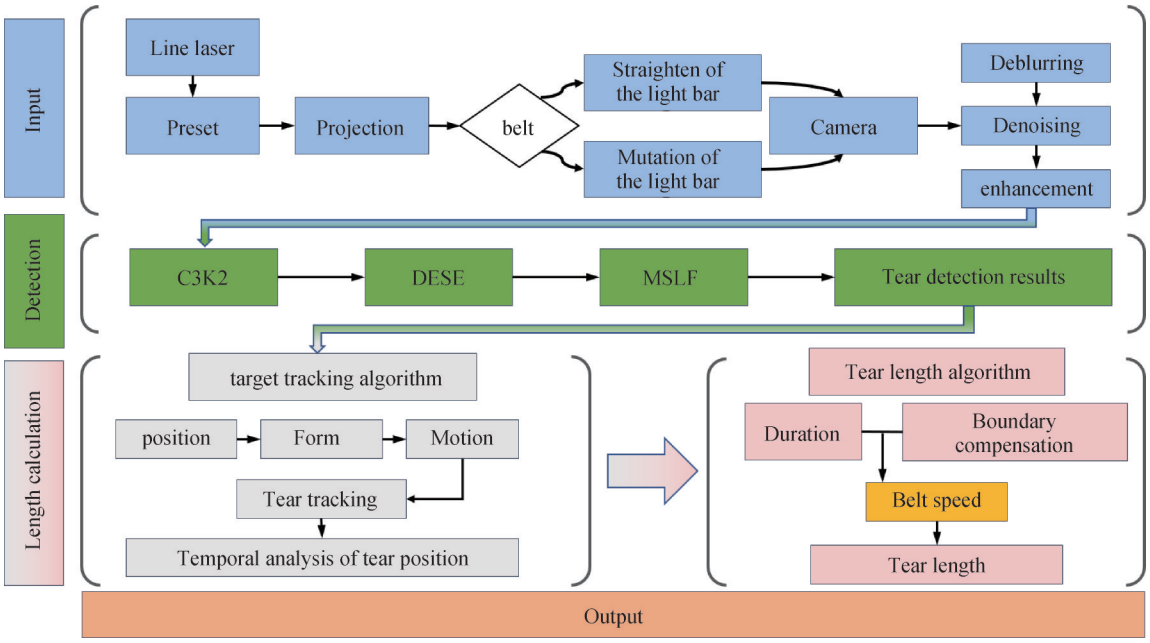

Fig.5 Workflow Diagram

However, when applied to conveyor belt tear detection in mining environments, small objects often experience feature loss after multiple downsampling iterations. The model also struggles with processing linear characteristics, brightness distribution, and multi-scale information within the laser stripes, limiting its ability to detect small targets and extract linear laser features effectively. To overcome these challenges, this study builds upon YOLOv11n by introducing the DESE dual-attention mechanism in the backbone network to enhance key feature representation. Additionally, a Multi-Scale Laser Feature Fusion (MSLF) module is designed to specifically address laser line feature extraction. The overall architecture of the improved model is shown in Fig. 6.

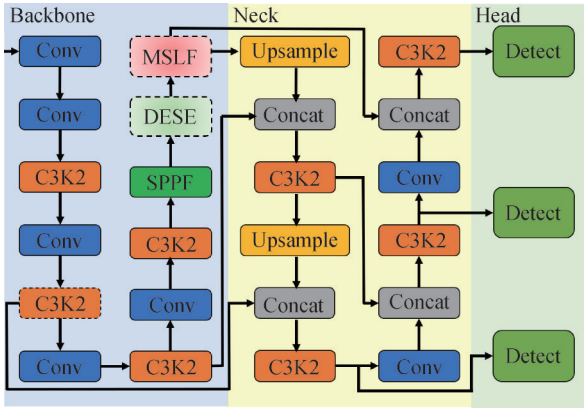

Fig.6 YOLO-DM Network Architecture Diagram

3.1 DESE Dual Attention Mechanism

Traditional attention mechanisms have limitations when applied to conveyor belt tear detection. SE attention models channel dependencies through fully connected layers, with computational complexity of $O(C^2/r)$ where C represents the number of channels and r is the reduction ratio. Although this global modeling approach captures long-range dependencies, it neglects local inter-channel correlations. ECA Attention, on the other hand, utilizes one- dimensional convolutions instead of fully connected layers, reducing computational complexity to $O(C\times k)$, where k is the convolut- ion kernel size. However, its local receptive field limits the ability to capture global features. To address these issues, this study introduces the DESE (Dual Enhanced Squeeze -and-Excitation) dual attention mechanism, which combines the strengths of both SE and ECA attention to enhance key features while maintaining computational efficiency.

The DESE mechanism adopts a dual-branch parallel processing architecture, as shown in Fig. 7. Laser

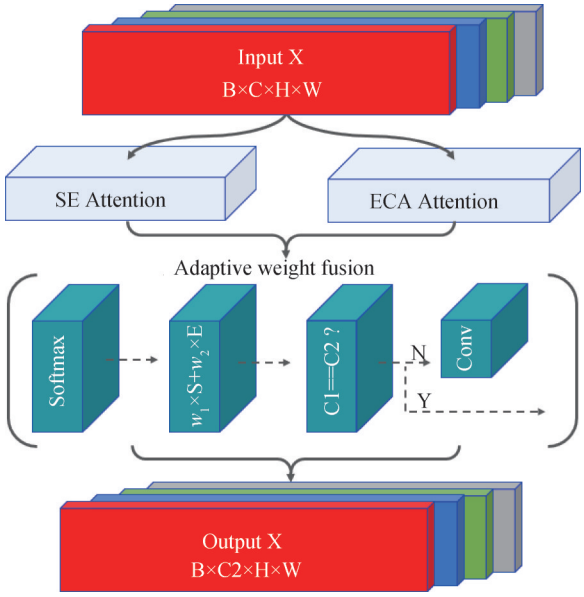

Fig.7 Structure diagram of DESE dual attention mechanism

intensity variations in tear regions exhibit localized and continuous characteristics, requiring enhanced feature representation along the channel dimension. In contrast, the geometric features of linear laser stripes have directional and continuous properties, necessitating strengthened feature correlation in the spatial dimension. The SE attention branch excels at global feature modeling, while the ECA attention branch provides computational efficiency and is well-suited for local feature processing.

The SE attention branch captures global contextual information and is well suited to modeling holistic semantics in tear regions. Specifically, spatial information is aggregated via global average pooling, after which a two-layer fully connected network learns inter-channel dependencies and produces channel attention weights. The SE-branch computation is formulated as:

$$y = \mathrm{GlobalAvgPool}(x) \tag{6}$$

$$s = \sigma\left(W_2 \cdot \mathrm{ReLU}\left(W_1 \cdot y\right)\right) \tag{7}$$

$$x_{SE} = x \odot s \tag{8}$$

where W_1 and W_2 denote the weight matrices of the two fully connected layers, $\sigma(\cdot)$ is the Sigmoid activation function, and $\odot$ represents element- wise multiplication.

The ECA attention branch promotes localized channel interaction using an efficient one-dimensional convolution, strengthening feature representation with low overhead. In this design, the fully connected operation is replaced with a 1D convolution, thereby substantially reducing the parameter count and computational cost. The convolution kernel size in ECA is determined by an adaptive mapping:

$$k = \psi(C) = \left| \frac{\log_2(C)}{\gamma} + \frac{b}{\gamma} \right|_{odd} \tag{9}$$

where C is the number of channels, γ and b are hyperparameters, and $|\cdot|_{odd}$ means taking the nearest odd number.

A key contribution of DESE lies in its adaptive fusion strategy, which dynamically balances the two branches according to the statistical characteristics of the input features. The fusion weights are produced by a lightweight gating network:

$$w = \mathrm{Softmax}\left(\mathrm{Linear}\left(\mathrm{GlobalAvgPool}(x)\right)\right) \tag{10}$$

$$x_{DESE} = w_1 \cdot x_{SE} + w_2 \cdot x_{ECA} \tag{11}$$

where w_1 and w_2 are the learned adaptive weights, satisfying $w_1 + w_2 = 1$.

3.2 *Multi-Scale Laser Line Feature Fusion (MSLF)*

Projected laser-stripe patterns on conveyor belts exhibit strong line-structure dominance and pronounced directionality, which differ markedly from the texture-rich statistics of natural images. Under normal operating conditions, the stripe remains continuous, maintains a stable principal orientation (typically orthogonal to belt motion), exhibits an approximately constant width, and shows relatively uniform intensity. Once a longitudinal tear occurs, the stripe commonly exhibits three representative mutations:

(1)Positional mutation: localized lateral displacement;

(2)Intensity mutation: brightness fluctuation induced by scattering from irregular tear edges;

(3)Morphological distortion: bending, localized orientation changes, or fragmentation due to surface inclination.

These cues are often subtle and can be attenuated during feature aggregation in conventional multi-scale fusion (neck) structures. To mitigate this issue, we introduce a Multi-Scale Laser Line Feature Fusion (MSLF) module (Fig. 8), which preserves stripe-related cues across multiple receptive fields and explicitly enhances direction-inconsistent regions that are more likely to correspond to tears.

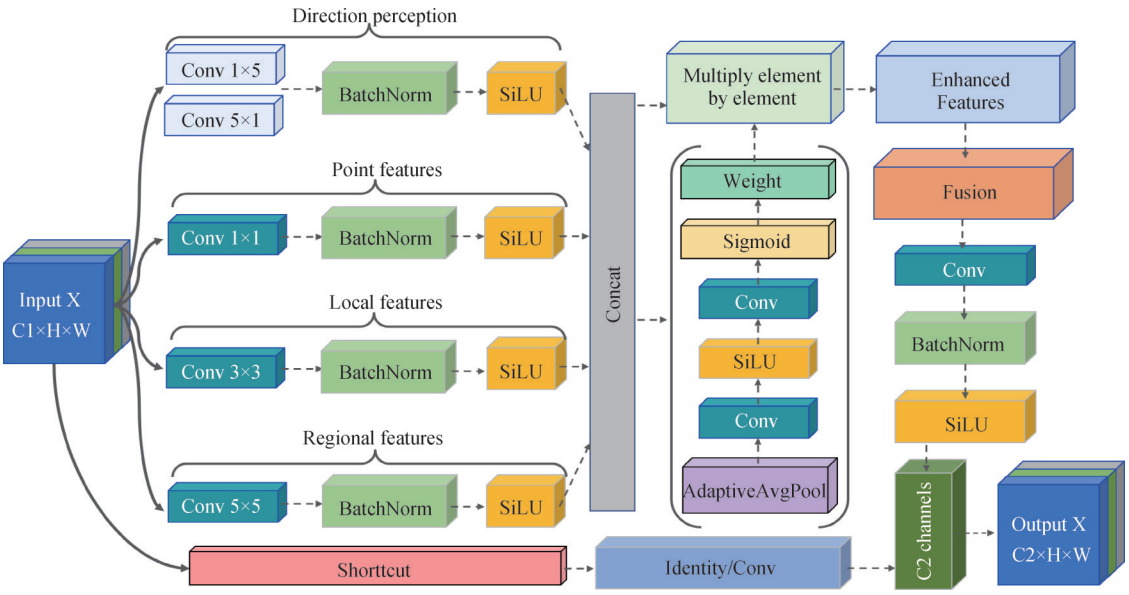

Fig.8 Structure diagram of the MSLF multi-scale laser line feature fusion module

3.2.1 Multi-scale feature extraction

Given an input feature map $X \in R^{C_1 \times H \times W}$, MSLF constructs four parallel branches (each followed by BatchNorm and SiLU) to extract laser-stripe features at different spatial scales:

Point branch (1×1 conv): captures point-wise intensity variation and channel-mixing cues, which are sensitive to small brightness changes and fine discontinuities.

Local branch (3×3 conv): captures short- range stripe deformation and local breaks.

Regional branch (5×5 conv): captures larger- scale bending and regional structural distortion.

Direction perception branch (asymmetric convs): employs anisotropic kernels to highlight direction-related variations of the stripe.

Denote the outputs of the four branches as $\{f_i\}_{i=1}^{4}$. They are concatenated along the channel dimension to form a unified representation:

$$F = \mathrm{Concat}(f_1, f_2, f_3, f_4) \tag{12}$$

This design maintains both fine-grained and large-context cues without forcing a single receptive field to explain all tear patterns.

3.2.2 Direction-aware Enhancement Mechanism

A key observation is that the projected stripe has a dominant direction in normal conditions, whereas tear regions introduce local direction inconsistency. Based on this property, MSLF introduces a direction-aware enhancement mechanism consisting of two steps: direction perception and consistency-guided reweighting.

(1) Direction perception via asymmetric convolution

To emphasize direction-dependent variations, MSLF adopts asymmetric kernels. When the stripe is approximately vertical, a (1×5) kernel is particularly effective at capturing lateral variations, which directly correspond to stripe offset and edge irregularities at tear sites. In practice (Fig. 8), both (1×5) and (5×1) kernels are used to maintain robustness when the stripe direction is not perfectly aligned with the image axes.

(2) Direction consistency estimation

Let *angle* be the local gradient direction angle (e.g., computed using Sobel gradients on the stripe intensity map or shallow features). Let $main_{direction}$ denote the dominant stripe direction, estimated from the stripe region (for example, by PCA over stripe pixels or by averaging local dominant orientations within a window). The direction *consistency* is defined as:

$$\mathrm{consistency} = \cos\left(\mathrm{angle} - \mathrm{main}_{\mathrm{direction}}\right) \tag{13}$$

This measure is close to 1 when local orientation agrees with the global stripe direction, and decreases as the local orientation deviates due to bending, scattering, or fragmentation.

(3) Consistency-guided reweighting and feature enhancement

To obtain a bounded and smooth attention weight, the cosine consistency is mapped by a Sigmoid function:

$$\mathrm{weights} = \sigma\left(5 \times \mathrm{consistency}\right) \tag{14}$$

where $\sigma(\cdot)$. denotes the Sigmoid function and the coefficient 5 controls the slope of the mapping. The resulting weight map $\mathrm{weights} \in (0, 1)$ is applied to the concatenated feature map via element-wise multiplication (corresponding to "Multiply element by element" in Fig. 8):

$$F_{enh} = F \odot \mathrm{weights} \tag{15}$$

where $\odot$ indicates element-wise multiplic- ation (broadcasted to channels). Intuitively, this operation suppresses direction-consistent regions that are less informative (normal stripe) and emphasizes direction-inconsistent regions where tear-induced stripe mutations are likely to appear.

3.2.3 Scale-adaptive Feature Selection and Fusion

Tears with different sizes and severities produce stripe mutations at different spatial extents. Therefore, MSLF introduces a scale- adaptive selection mechanism to dynamically adjust the contributions of the four branches. Specifically, a compact descriptor is extracted from each branch by global average pooling (GAP), concatenated, and passed through a lightweight linear layer followed by Softmax to obtain scale weights:

$$w_{scale} = \mathrm{Softmax}\left(\mathrm{Linear}\left(\mathrm{Concat}\left(\left[\mathrm{GAP}(f_i)\right]_{i=1}^{4}\right)\right)\right) \tag{16}$$

where, $w_{scale} = [w_1, w_2, w_3, w_4]$ satisfies $\sum_{i=1}^{4} w_i = 1$. The weighted fusion (implemented by the "Fusion→Conv→ BatchNorm →SiLU" block in Fig. 8) aggregates multi-scale information, while a shortcut path ("Identity/Conv") preserves stable information and facilitates optimization. Consequently, the MSLF module can (i) retain stripe cues across receptive fields, (ii) explicitly enhance direction-inconsistent mutation regions, and (iii) adaptively fuse multi-scale features according to tear morphology, improving robustness under dust, blur, and illumination fluctuations encountered in mining environments.

3.3 Dataset Construction and Annotation

To train and evaluate YOLO-DM, a dedicated conveyor-belt longitudinal-tear dataset was constructed. Owing to safety constraints and the requirement for uninterrupted operation in underground mines, data were collected in a laboratory environment rather than at industrial sites. This choice is justified by the following considerations.

Targeted detection objective. The proposed method focuses on identifying geometric discontinuities in projected laser stripes (positional shifts, intensity variations, and morphological distortions). These cues are primarily governed by tear geometry, while environmental factors mainly affect visibility and imaging quality. By controlling tear types and

configurations in the laboratory, high-quality annotations can be obtained, providing clear feature benchmarks for model learning.

Coverage of multiple tear types. The dataset includes minor tears (5 – 20 cm in length, 1 – 3 mm in depth), moderate tears (20 – 80 cm, 3 – 10 mm), severe tears (>80 cm, >10 mm), and multiple-tear scenarios. To narrow the gap between laboratory conditions and real mining environments, a multi-dimensional augmentation strategy was adopted to enhance diversity and generalization. Three augmentation operations were applied: dust simulation, motion blur, and lens distortion.

(1) Dust simulation: Perlin noise combined with grayscale haze is overlaid to emulate coal-dust conditions. Three dust levels—light, moderate, and heavy—are generated using transparency parameters in the range 0.15 – 0.45, which reduce contrast and soften laser-stripe edges, thereby improving robustness under low-visibility conditions;

(2) Motion blur: A directional blur kernel aligned with belt motion is applied, with kernel sizes ranging from 5 to 15 pixels, reproducing trailing artifacts during high-speed operation (belt speed > 2 m/s);

(3) Lens distortion: A radial distortion model is used to mimic wide-angle lens deformation, covering pincushion distortion ($k_1 \in [-0.15, -0.05]$) and barrel distortion ($k_1 \in [0.05, 0.15]$), improving geometric-feature robustness under optical distortion.

Using the above augmentation pipeline, the original 1, 200 images were expanded to 3, 000 high-quality annotated images. Examples are shown in Fig. 9.

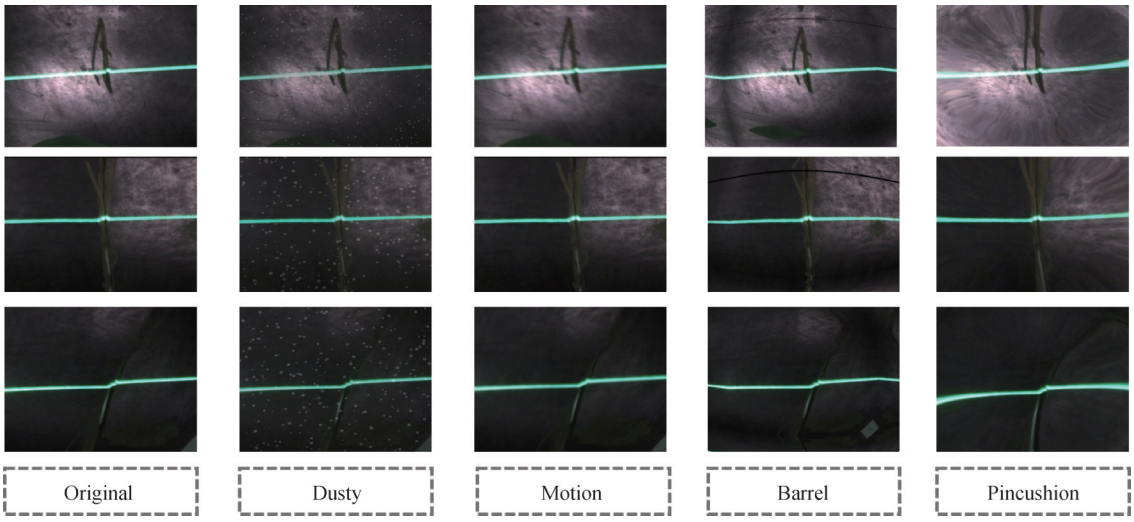

Fig.9 Expansion Example of Longitudinal Tearing Dataset

Tear regions were annotated in LabelImg using rectangular bounding boxes under strict labeling guidelines: (i) each box must fully cover the visible tear area, including the tear body and prominent edge deformation; (ii) pixel-level precision is required along the longitudinal axis; (iii) when the laser stripe is interrupted, the box must include the adjacent stripe segments before and after the interruption; and (iv) multiple tears must be labeled separately to avoid merged annotations. Criteria for identifying laser-stripe discontinuities are defined as follows: lateral stripe shifts exceeding 3 pixels are treated as valid mutations; intensity changes exceeding 30% relative to the baseline are considered anomalies; and continuous stripe interruptions longer than 10 pixels are regarded as tear features.

The final dataset contains 3, 000 annotated images, split into training/validation/test sets at a ratio of 7: 2: 1. The training set includes 2, 100 images (1, 890 tear samples), the validation set contains 450 images (405 tear samples), and the test set contains 450 images (391 tear samples). Tear categories are distributed as minor (1, 075), moderate (940), and severe (671), ensuring a balanced and representative dataset.

4 Tear Length Calculation Algorithm

The precise measurement of tear length serves as a crucial basis for conveyor belt damage assessment and maintenance decisions. Traditional manual measurement methods are inefficient and pose safety hazards, while image-based approaches often suffer from reduced accuracy due to factors such as viewing angle and lighting conditions. This chapter proposes an automated tear length calculation algorithm integrating target tracking with belt speed data. By performing temporal localisation and trajectory tracking of tear targets across consecutive frames, combined with conveyor belt operational velocity information, it achieves high-precision real-time tear length measurement.

4.1 Tracking of Tear Positions

4.1.1 Tearing Target Association Across Consecutive Frames

Accurate association of tear instances across frames is fundamental to reliable length estimation. Because the

belt moves continuously, the same tear can shift noticeably between adjacent frames; robust data association is therefore required to preserve track continuity.

To improve inter-frame association accuracy, we adopt a multi-cue fusion strategy that jointly considers geometric proximity, appearance affinity, and motion consistency:

$$S_{total}(i,j) = \alpha S_{pos}(i,j) + \beta S_{app}(i,j) + \gamma S_{motion}(i,j) \quad (17)$$

Where $S_{total}(i,j)$ is the comprehensive similarity between the detected target i in the t frame and the detected target j in the $(t+1)$ frame; $S_{pos}(i,j)$, $S_{app}(i,j)$, and $S_{motion}(i,j)$ are the position similarity, appearance similarity, and motion similarity respectively; α, β, and γ are the corresponding weight coefficients, and $\alpha+\beta+\gamma=1$.

The position similarity is calculated based on the overlap degree of the detection boxes:

$$S_{pos}(i,j) = \frac{\text{Area}(B_i \cap B_j)}{\text{Area}(B_i \cup B_j)} \quad (18)$$

Where B_i and B_j are the detection boxes of the corresponding targets in the two frames respectively; $\text{Area}(\cdot)$ represents the area calculation function.

Appearance similarity is measured by the cosine distance of feature vectors:

$$S_{app}(i,j) = \frac{F_i \cdot F_j}{|F_i| \cdot |F_j|} \quad (19)$$

Where F_i and F_j are the appearance feature vectors of the two detections. In this work, they are 512-dimensional embeddings extracted from the YOLOv11n feature layer.

Motion similarity. To enhance robustness under belt motion, a Kalman filter is used to predict the tear state. Considering the approximately constant belt speed, we adopt a constant-velocity state model:

$$\mathbf{x}_{k+1} = \mathbf{F}\mathbf{x}_k + \mathbf{w}_k \quad (20)$$

the state vector $x_k = \left[x_k, y_k, \dot{x}_k, \dot{y}_k\right]^T$ includes the target's position coordinates (x_k, y_k) and velocity components $(\dot{x}_k, \dot{y}_k)$; The state transition matrix **F** and process noise $\mathbf{w}_k$ are defined as:

$$\mathbf{F} = \begin{bmatrix} 1 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & \Delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \ \mathbf{w}_k \sim N(0, Q) \quad (21)$$

The motion similarity is calculated by the reciprocal of the distance between the predicted position and the actual detected position:

$$S_{motion}(i,j) = \frac{1}{1 + |\mathbf{P}_{pred,i} - \mathbf{P}_{det,j}|_2} \quad (22)$$

where $\mathbf{p}_{pred,i}$ is the position of target *i* in the next frame predicted by the Kalman filter; $\mathbf{p}_{det,j}$ is the actual detected position of target*j*.

The weights α, β, γ in Eq. (17) reflect the reliability of three cues (position overlap, appearance embedding similarity, and motion consistency). In mining scenarios, cue reliability changes with dust, blur, and belt speed. Therefore, instead of using fixed empirical values, we set the weights using a reliability-aware fusion rule derived from minimum-variance linear estimation:

$$\alpha = \frac{\sigma_{pos}^{-2}}{\sigma_{pos}^{-2} + \sigma_{app}^{-2} + \sigma_{motion}^{-2}} \quad (23)$$

$$\beta = \frac{\sigma_{app}^{-2}}{\sigma_{pos}^{-2} + \sigma_{app}^{-2} + \sigma_{motion}^{-2}} \quad (24)$$

$$\gamma = \frac{\sigma_{motion}^{-2}}{\sigma_{pos}^{-2} + \sigma_{app}^{-2} + \sigma_{motion}^{-2}} \quad (25)$$

where $\sigma_{pos}, \sigma_{app}, \sigma_{motion}$denote the empirical uncertainty (standard deviation) of each cue. These uncertainties can be estimated online using a short sliding window of matched tracks:

(1)$\sigma_{pos}$from the variance of IoU between predicted and matched boxes;

(2)$\sigma_{app}$from the variance of cosine similarity of the 512-d embeddings across consecutive matched frames;

(3)$\sigma_{motion}$from the variance of Kalman residuals (innovation) between predicted and observed positions.

This strategy automatically down-weights the cue that becomes unstable (e.g., appearance under heavy dust, motion under high speed and blur).

4.1.2 Tear Identity Discrimination Mechanism

To determine whether detections correspond to the same physical tear, a progressive identity discrimination scheme is employed based on similarity, temporal continuity, and spatial plausibility.

First, a comprehensive similarity threshold θ_{sim} is set. When $S_{total}(i,j) > \theta_{sim}$, targets *i* and *j* are considered to be the same tear. Based on experimental statistics, θ_{sim} is set to 0.6.

Second, a temporal continuity constraint is introduced by limiting the maximum permissible missing duration T_{max}:

$$T_{missing} = (t_{current} - t_{last\_detected}) \times \frac{1}{fps} \quad (26)$$

If $T_{missing} > T_{max}$ the tear is considered to have left the detection region and the track is terminated. This constraint mitigates track drift caused by prolonged detection gaps while maintaining computational efficiency.

Third, to further reduce cross-tear mismatches, a spatial continuity constraint is imposed based on expected belt motion. A position-jump threshold is defined as:

$$d_{max} = v_{belt} \times \Delta t \times k_{margin} \quad (27)$$

Where v_{belt} is the conveyor belt speed; Δt is the inter-frame time interval; k_{margin} is the safety factor (set to 1.5). This constraint ensures the spatial rationality of target association and effectively eliminates abnormal

associations caused by environmental interference or detection noise.

Through the above three-level progressive constraints, tear identity discrimination is substantially improved, providing a reliable basis for subsequent length estimation.

4.2 Tear Length Calculation Model

Tear length is estimated from the residence time of the tear within the detection region together with belt speed. As a tear traverses the field of view with the moving belt, the presence duration in the image satisfies:

$$L_{tear} = v_{belt} \times \Delta t_{presence} \tag{28}$$

where L_{tear} is the tear length; v_{belt} is the running speed of the conveyor belt; $\Delta t_{presence}$ is the duration of the tear in the detection area. However, calculations based solely on simple time differences often have errors, because the first and last detection times of the tear cannot fully represent the real times when it enters and leaves the detection area. To improve measurement accuracy, it is necessary to refine the modeling of the tear duration, considering the boundary effects when the tear enters and exits the detection area.

Therefore, the accurate calculation of the tear duration should include boundary time compensation:

$$\Delta t_{presence} = t_{end} + t_{start} + \Delta t_{entry} + \Delta t_{exit} \tag{29}$$

where t_{start}: the time when the tear is first detected; t_{end}: the time when the tear is last detected; Δt_{entry}: the time compensation from when the tear enters the detection area to when it is first detected; Δt_{exit}: the time compensation from when the tear is last detected to when it exits the detection area.

Considering the uniform linear motion characteristics of the tear with the conveyor belt, the time compensation terms have a direct proportional relationship with the spatial distance:

$$\Delta t_{entry} = \frac{d_{entry}}{v_{belt}}, \Delta t_{exit} = \frac{d_{exit}}{v_{belt}} \tag{30}$$

Where d_{entry} and d_{exit} are the distance compensations for the tear entering and exiting the detection area respectively.

With the above tracking-and-compensation design, the proposed algorithm enables quantitative evaluation of tear severity and provides accurate, actionable measurements to support maintenance decisions, thereby improving the intelligence and reliability of safety monitoring for mining conveyor belts.

5 Experimental Results and Analysis

To validate the effectiveness of the proposed line-laser- and YOLO-DM-based longitudinal tear detection system for mining conveyor belts, this section comprehensively evaluates detection accuracy, real-time performance, and tear-length estimation precision through experimental platform construction, comparative studies, and ablation experiments.

5.1 Experimental Platform and Model Performance Evaluation

(1) Experimental Environment Setup

Model training and testing were conducted on a high-performance industrial computer (Fig. 4(d)). The implementation was based on PyTorch 1.13.0, using Python 3.9.10 and CUDA 11.7. The remaining hardware components are shown in Fig.4.

To ensure fairness and reproducibility, all models were trained under identical hyperpara- meter settings: the input resolution was fixed at 640×640, the number of training epochs was set to 300, and the batch size was 16. AdamW was used as the optimizer with an initial learning rate of 0.001, and a cosine annealing schedule was applied.

(2) Evaluation Metrics

Detection performance was assessed using Precision (P), Recall (R), Frames Per Second (FPS), and Mean Average Precision (mAP). In addition, to quantify computational efficiency and deployment cost, we also report the model computational complexity in GFLOPs and the parameter count (Param/10^6), which reflect arithmetic cost and memory footprint, respective- ly. The corresponding definitions are given by:

$$P = \frac{TP}{TP + FP} \times 100\% \tag{31}$$

$$R = \frac{TP}{TP + FN} \times 100\% \tag{32}$$

$$FPS = \frac{1000}{speed} \tag{33}$$

$$mAP = \frac{1}{n}\sum_{i=1}^{n}\int_{0}^{1} P_i(R)\,dR \times 100\% \tag{34}$$

Where TP denotes the number of positive samples correctly identified as positive; FP denotes the number of negative samples incorrectly identified as positive; FN denotes the number of positive samples incorrectly identified as negative; speed denotes the sum of model pre-processing, post-processing, and inference time; n denotes the number of categories.

(3) Model Performance Evaluation

To benchmark detection performance, YOLO-DM was compared with SSD, Faster R-CNN, YOLOv7, YOLOv8, YOLOv10, and YOLOv11n on the constructed conveyor-belt tear dataset. Training curves are presented in Fig.10, where YOLO-DM achieves the highest mAP@0.5 and mAP@0.5:0.95 across the evaluated models.

All methods were further evaluated on the test set, with the quantitative results summarized in Table 1. YOLO-DM achieves an mAP@0.5 of 98.7%, representing a 9.2 percentage-point improvement over the baseline YOLOv11n. Its mAP@0.5 – 0.95 reaches 71.2% (+8.9 percentage points). Precision and recall are 96.4%

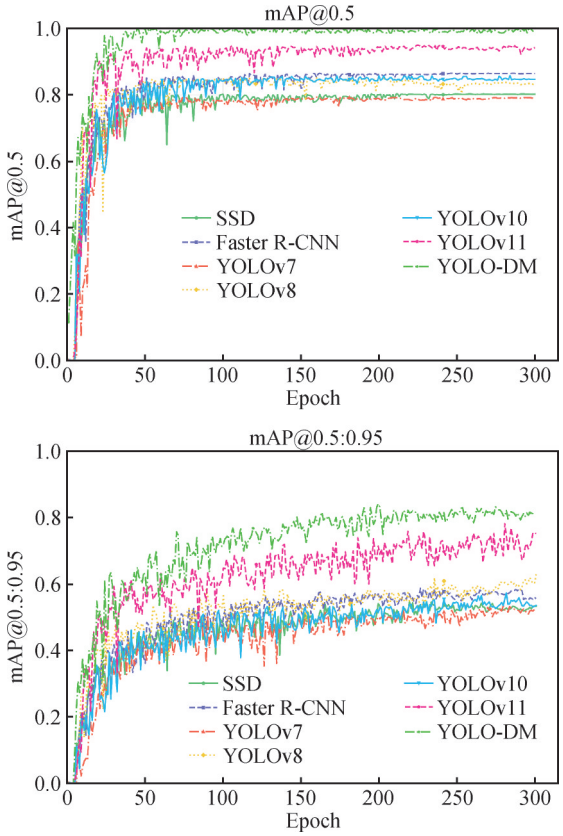

Fig.10 Performance comparison of different algorithm models

and 98.1%, respectively, indicating a favorable balance between detection accuracy and sensitivity.

As evidenced in Table 1, YOLO-DM consistently outperforms mainstream detectors across multiple key metrics. In terms of accuracy, its mAP@0.5 (98.7%) exceeds YOLOv11n by 9.2 percentage points and YOLOv8n by 18.3 percentage points, while its mAP@0.5 –0.95 (71.2%) substantially surpasses traditional methods such as SSD and Faster R-CNN. In terms of efficiency, YOLO-DM achieves a per-frame inference time of 17.2 ms, corresponding to 58.1 FPS, which satisfies real- time deployment requirements. Specifically, feature extraction requires 14.4 ms, detection- head inference takes 1.9 ms, and post-processing costs 0.9 ms.

To further reflect deployment cost, Table 1 also reports GFLOPs and parameters. Compared with two-stage Faster R-CNN (370.1 GFLOPs, 63.20M parameters), YOLO-DM requires 14.70 GFLOPs and 8.67M parameters, indicating a much lower computation and memory footprint while achieving substantially higher accuracy and real-time speed. Compared with the baseline YOLOv11n (8.30 GFLOPs, 8.88M), YOLO-DM attains a marked accuracy gain with a moderate increase in GFLOPs, suggesting that the proposed feature enhancement improves recognition capability at an acceptable computational cost for real-time monitoring.

Table 1 Performance comparison of models on the test set

Model	P/%	R/%	mAP@0.5/%	mAP@0.5:0.95/%	FPS	GFLOPs	Param/10^6
SSD	83.6	74.8	78.8	49.5	27.5	63.20	12.30
Faster R-CNN	87.3	76.6	83.5	53.4	11.2	370.1	63.20
YOLOv7n	82.9	81.2	77.2	48.3	53.4	13.2	9.10
YOLOv8n	81.8	83.4	80.4	51.3	54.6	8.50	9.70
YOLOv10n	84.3	82.7	79.4	49.8	51.2	9.21	8.92
YOLOv11n	87.9	91.2	89.5	62.3	75.7	8.30	8.88
YOLO-DM	96.4	98.1	98.7	71.2	58.1	14.70	8.67

Overall, YOLO-DM not only delivers superior detection accuracy compared with representative baselines, but also maintains strong real-time throughput and computational efficiency, supporting its deployment on edge devices.

5.2 Ablation Study

To quantify the contribution of each proposed component, three ablation settings were designed to evaluate the DESE dual-attention mechanism, the MSLF module, and their combined effect. As shown in Fig. 11, YOLO-DM exhibits a faster mAP increase during training than the other configurations and maintains a consistently higher mAP throughout the training process.

Detailed results are reported in Table 2. Incorporating DESE improves mAP@0.5 and mAP@0.5–0.95 by 1.6 and 1.5 percentage points, respectively. Notably, the detection accuracy for micro-tears increases by 4.2 percentage points, confirming the benefit of dual-branch attention in enhancing weak tear cues. Meanwhile, inference speed increases slightly by 0.4 FPS, indicating that DESE is lightweight and efficient. From the complexity perspective, +DESE also reduces computational cost (7.51 GFLOPs) and model size (7.12M parameters) compared with the YOLOv11 baseline (8.30 GFLOPs, 8.88M), suggesting that the adopted attention design enhances representation while remaining deployment-friendly.

After introducing MSLF, recall improves by 5.9 percentage points, and mAP@0.5 increases by 6.8 percentage points, demonstrating that the module effectively strengthens laser-stripe anomaly features. This

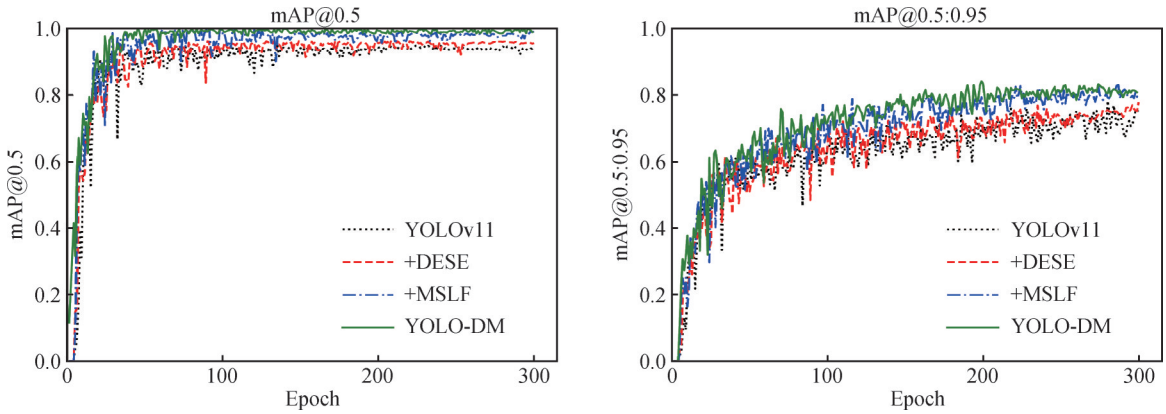

Fig.11 Performance comparison of different module algorithm models

Table 2 Ablation study results

Model	DESE	MSLF	P/%	R/%	mAP@0.5/%	mAP@0.5:0.95/%	FPS	GFLOPs	Param/10^6
YOLOv11	-	-	87.9	91.2	89.5	62.3	75.7	8.30	8.88
+DESE	√	-	89.2	97.9	91.1	63.8	76.1	7.51	7.12
+MSLF	-	√	94.5	97.1	96.3	68.9	60.7	13.56	8.16
YOLO-DM	√	√	96.4	98.1	98.7	71.2	58.1	14.70	8.67

gain comes with an expected increase in computation (13.56 GFLOPs) while keeping parameters moderate (8.16M), because MSLF uses parallel multi-scale convolutions to preserve stripe cues at different receptive fields.

By combining DESE and MSLF, the final YOLO-DM model achieves substantial accuracy gains while maintaining real-time speed, validating both the effectiveness of the proposed enhancements and their favorable synergy. In addition, YOLO-DM maintains a compact parameter scale (8.67M) while increasing GFLOPs to 14.70, which is a deliberate accuracy-efficiency trade-off that remains suitable for real-time industrial monitoring.

Although the dataset was collected in a laboratory environment, the proposed design is tailored to the intrinsic structured-light characteristics that are common across mining sites—i. e., tear-induced stripe positional shifts, intensity fluctuations, and morphological distortions. The DESE module strengthens weak cues for early-stage tears, while MSLF provides multi-scale receptive fields and direction-aware enhancement, improving robustness when belt speed, dust level, blur, or laser orientation varies. Together with the adopted augmentation strategy (dust simulation, motion blur, lens distortion), YOLO-DM is expected to generalize better to different conveyor operating conditions and imaging setups. Moreover, the framework remains modular: the detector can be retrained with site-specific data, and the same tracking-and-speed integration logic can be reused across different belt widths and speeds, facilitating practical adaptation to diverse application scenarios.

5.3 Verification of Length Measurement Accuracy

To verify the accuracy of the tear length calculation algorithm, artificial tear samples of varying lengths were created for testing. The test samples, shown in Fig. 12, include five length specifications: 5 cm, 10 cm, 20 cm, 40 cm, and 80 cm.

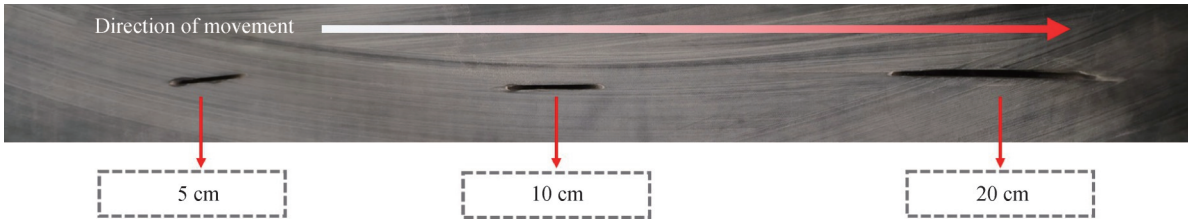

Fig.12 Partial view of the test sample

Experiments were conducted at a fixed belt speed of 1 m/s. Comparing manual measurements with the proposed estimator shows an average absolute error of 1.2 cm, corresponding to an overall measurement accuracy of 95.8%. For tears longer than 20 cm, the accuracy increases to 97.3%, whereas for minor tears of 5

– 20 cm, it decreases to 91.5%. For clarity and comparability, the measurement accuracy is defined as:

$$\mathrm{Acc} = 1 - \frac{\left|L_{est} - L_{gt}\right|}{L_{gt}} \times 100\%. \tag{35}$$

The measurement accuracy under different belt speeds is presented in Fig. 13.

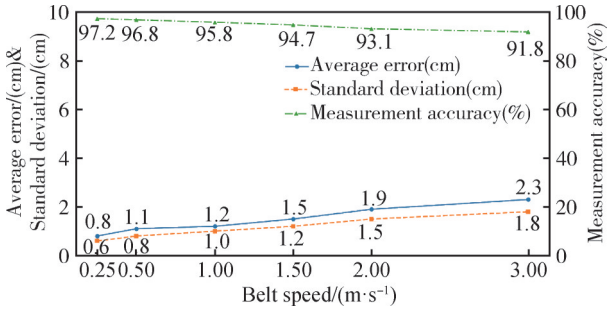

Fig.13 Measurement accuracy under different conveyor-belt speeds

As belt speed increases, measurement accuracy exhibits a slight downward trend. Specifically, when belt speed increases from 0.25 m/s to 3.0 m/s, the average absolute error increases from 0.8 cm to 2.3 cm, the standard deviation increases from 0.6 cm to 1.8 cm, and the measurement accuracy decreases from 97.2% to 91.8% (Fig. 13). Nevertheless, within the commonly used operating range of 1–2 m/s, the accuracy remains above 93.1%, satisfying practical engineering requirements. The degradation at higher speeds is mainly attributed to increased tracking errors caused by motion blur and larger inter-frame displacement.

A statistical analysis based on 20 repeated measurements indicates that the estimation errors approximately follow a normal distribution. The mean absolute error is 1.2 cm with a standard deviation of 0.95 cm, and the errors within the 95% confidence interval lie within ±2.1 cm. Error-source analysis suggests that detection accuracy accounts for approximately 45% of the total error, belt-speed estimation contributes about 30%, and target tracking contributes around 25%. These results indicate that further improvements in measurement precision are achievable by enhancing detection robustness and refining the tracking strategy.

To further demonstrate the advantages of the proposed length measurement method, we compare it with representative studies. Xu et al.[16] reported a multi-linear-laser-based longitudinal-tear system with a mean length estimation error of 0.06 m (6 cm), which typically requires multiple laser lines and additional calibration/maintenance, increasing deployment complexity and hardware cost. In contrast, our method achieves a mean absolute error of 1.2 cm (0.012 m) at 1 m/s using a single line laser + single camera, thereby reducing sensing hardware complexity while improving accuracy. In addition, You et al. [31] reported a fully convolutional neural network (FCN)-based approach with a maximum relative error of 13.04% for physical-size measurement. Benefiting from the boundary-compensated duration model and belt-speed integration, our method maintains > 93.1% accuracy in the common speed range of 1–2 m/s, and remains 91.8% even at 3 m/s (Fig. 13). From an efficiency perspective, the proposed length estimation only relies on online association/tracking of detection outputs and belt-speed integration, introducing minimal extra computation beyond the detector, and is therefore suitable for real-time deployment while keeping system cost low due to the compact sensing configuration.

Table 3 Comparison of tear-length/size measurement performance and deployment implications

Method	Sensing configuration	Measurement target	Reported error (from literature)	Online/real-time implication	Cost & deployment complexity
Xu et al. [16] (2021)	Multi-linear lasers + camera	Damage length	Mean error 0.06 m	Multi-line processing; calibration needed	Higher hardware + calibration/ maintenance cost
You et al. [31] (2022)	Camera (FCN-based)	Physical size of tear	Max relative error 13.04%	Pixel-level estimation; compute-heavy tendency	Higher compute demand for edge deployment
This work (YOLO-DM + tracking)	Single line laser + single camera	Tear length	Mean abs. error 1.2 cm (1 m/s); accuracy >93.1% (1–2 m/s)	Tracking + speed integration; minimal extra overhead beyond detector	Compact hardware; lower deployment complexity

5.4 Visual Analysis of Detection Results

This subsection provides an intuitive qualitative assessment of YOLO-DM on line-laser imagery, and further discusses long-term operational stability, reliability considerations, and practical deployment scalability. In addition to quantitative metrics (Tables 1–2) and length-estimation accuracy (Section 5.3), qualitative visualization is important for understanding how the model responds to typical laser-stripe mutations caused by tearing, including stripe displacement, intensity fluctuation, and partial interruption. For this purpose, we created the test conditions shown in Fig.14.

This involved producing three sets of quantitatively measured tears at lengths of 10 cm, 25 cm, and 35 cm, along with additional tears of arbitrary lengths. To

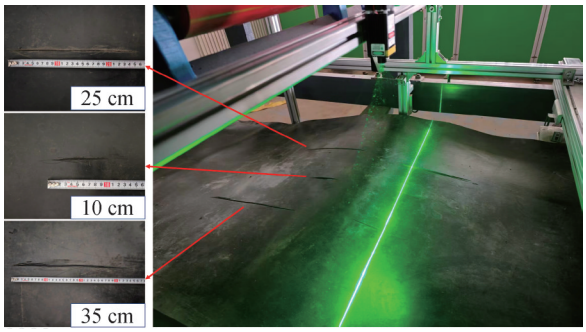

Fig. 14 Representative conveyor belt test samples, including predefined tear lengths (10 cm, 25 cm, and 35 cm), arbitrary tear lengths, and surface-abraded belts used for qualitative evaluation.

evaluate the model's generalization capability, two tear conveyor belts were fabricated, with one incorporating damage such as abrasion.

5.4.1 Qualitative Visualization of Detection and Online Length Output

Fig. 15 shows representative detection visualizations produced by the proposed system in real time. Each frame overlays: (1) detected tear bounding boxes; (2) the confidence score of each detection (label "tear" followed by confidence); and (3) the number of detected tears and their estimated tear lengths displayed at the upper-left corner. This visualization directly demonstrates the end-to-end pipeline "detection → tracking → length estimation," where the estimated length is obtained from the spatiotemporal tracking-based duration model with boundary compensation and belt-speed integration.

Several observations can be made from the examples in Fig. 15. First, the system can detect multiple tears simultaneously in a single frame, and each tear is assigned an individual length estimate, supporting multi-defect monitoring. Second, both short/incipient tears and long, clearly deformed tears are captured: the visualizations include cases with very small estimated length and weaker confidence (reflecting subtle stripe discontinuities) as well as cases with strong stripe distortion and high confidence. Third, the detections are spatially aligned with the laser-stripe anomaly regions, indicating that the proposed DESE attention and MSLF stripe-feature fusion effectively highlight stripe mutations that correspond to tearing.

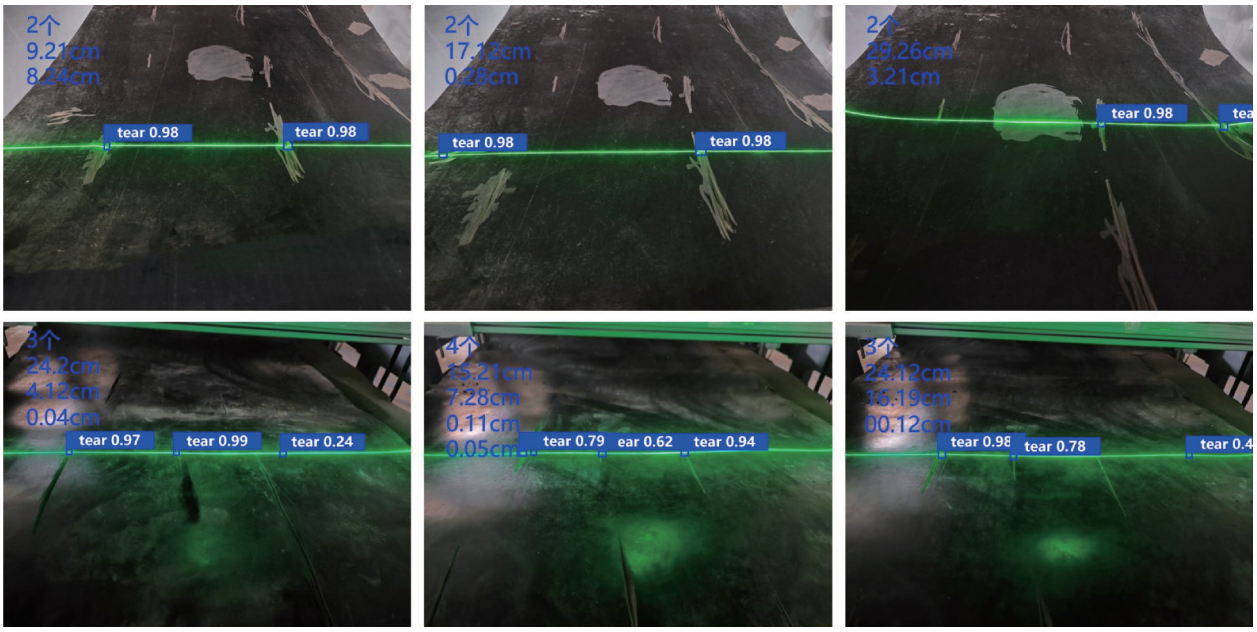

Fig.15 Qualitative detection visualization examples.

The upper-left text indicates the detected tear count and the estimated tear lengths; blue boxes show detected tear regions with confidence scores.

5.4.2 Practical Operation Discussion: Stability, Reliability, and Scalability

To reflect practical mining deployment, we summarize typical operation behavior and explain how the proposed design supports long-term stability and system scalability. In representative monitoring sequences, the system can handle: (1) multiple-tear scenarios, where each tear is tracked as an independent instance and accumulates its own presence duration for length estimation; (2) incipient tears, where the stripe mutation is weak—here the tracking and boundary-compensation strategy helps stabilize the final length output across frames rather than relying on a single image; and (3) large tears, where stripe bending or partial interruption occurs—MSLF enhances direction-inconsistent regions and preserves multi-scale cues, improving detection robustness.

For long-term stability and reliability, the system benefits from both hardware and algorithm design. The compact sensing configuration (single line laser + single

camera) reduces calibration burden and failure points compared with multi-line or multi-sensor solutions, which is advantageous for continuous operation in dusty and vibrational environments. On the algorithm side, DESE strengthens weak tear features, and MSLF provides direction-aware enhancement and scale-adaptive fusion, improving tolerance to dust, blur, and mild stripe-direction deviations. The length estimator further improves robustness through multi-cue association and reliability-aware fusion, helping maintain track continuity under transient detection degradation. For sustained deployment, it is recommended to monitor operational indicators such as false-alarm rate per hour, performance drift over time, stripe-quality, and throughput stability under continuous load.

Regarding scalability and integration, the proposed framework is modular and can be extended to wider belts by deploying multiple camera–laser units across the belt width while keeping each unit lightweight and independently maintainable; the detection, tracking, length- estimation logic remains unchanged per unit. The outputs can be integrated into existing PLC/SCADA workflows through standard industrial interfaces, enabling linkage with belt-speed signals, alarm logic, and maintenance logging. This provides a practical path from laboratory validation to real-world deployment and continuous operation.

6 Conclusions and Future Work

This study presents a real-time longitudinal tear monitoring framework that integrates single-line laser structured illumination, an enhanced YOLOv11 detector (YOLO-DM), and an online tear-length estimator. By actively projecting a laser stripe, the system converts subtle three-dimensional surface deformations into measurable stripe perturbations, thereby reducing dependence on ambient illumination relative to passive vision-based approaches. On this basis, YOLO-DM incorporates a dual-attention mechanism (DESE) and a multi-scale laser-stripe feature fusion module (MSLF) to tailor feature learning to the geometric and photometric characteristics of laser stripes. This design improves sensitivity to incipient tears while preserving real-time throughput. Experiments show that the proposed method achieves 98.7% mAP at 58.1 FPS, and the tear-length estimation reaches 95.8% accuracy, satisfying the real-time monitoring requirements in mining environments.

Beyond binary tear detection, the proposed framework enables online length quantification by integrating spatiotemporal target tracking with belt-speed integration and boundary compensation, providing actionable information for maintenance decision-making. The hardware configuration remains compact (single camera + single-line laser), supporting continuous in-situ monitoring, early warning, and quantitative assessment. Compared with contact-based devices, the proposed solution avoids mechanical wear, installation constraints, and frequent false alarms; compared with multi-laser or thermal-imaging systems, it offers improved portability and lower deployment complexity.

Future work will focus on: (1) expanding field datasets and addressing domain shift via adaptive/ continual learning to improve generalization under varying dust and vibration conditions; (2) deploying the model on embedded edge hardware through quantization and pruning while maintaining accuracy; (3) extending the evaluation from length to richer damage descriptors (e. g., width/depth proxies or severity grading) by incorporating geometric constraints and temporal consistency; (4) exploring multi-sensor fusion (e. g., vibration or acoustic signals) to enable earlier pre-tear warning and improve reliability; and (5) refining belt-speed acquisition and calibration to further reduce residual length-estimation errors.

Author Contribution:

Hao Hongtao: Project administration, Conceptualisation, Data curation, Formal analysis; Mai Xuewu: Methodology, Writing - original draft, Validation; Liu Chenhe: Visualisation, Writing - review & editing; Teng Pengfei: Investigation, Project administration; Ma Xiaodong: Writing - review & editing; Feng Baozhong: Funding acquisition.

Foundation Information:

This research was funded by the Key Research and Development Plan of Ningxia Hui Autonomous Region (2024BEE02029)

Data Availability:

The authors declare that the main data supporting the findings of this study are available within the paper and its Supplementary Information files.

Conflicts of Interest:

The authors declare no competing interests.

Dates:

Received 17 November 2025; Accepted 15 May 2026; Published online 10 July 2026

References

[1] Zimroz R., & Król R. (**2009**). Failure analysis of belt conveyor systems for condition monitoring purposes. *Mining Science* , 128(36), 255.

[2] He D., Pang Y., & Lodewijks G. (**2017**). Green operations of belt conveyors by means of speed control. *Applied Energy* , 188, 330-341.

[3] Grincova A., Andrejiova M., Marasova D., & Khouri S. (**2019**). Measurement and determination of the absorbed impact energy for conveyor belts of various structures under impact loading. *Measurement* , 131, 362-371.

[4] Zakharov A., Geike B., Grigoryev A., & Zakharova A. (**2020**). *Analysis of devices to detect longitudinal tear on conveyor belts* . In E3S Web of Conferences (Vol. 174, p. 03006). EDP Sciences.

[5] Marasová D., Ambriško Ľ., Andrejiová M., & Grinčová A. (**2017**). Examination of the process of damaging the top covering layer of a conveyor belt applying the FEM. *Measurement* , 112, 47-52.

[6] Fedorko G., Molnar V., Marasova D., et al. (**2014**). Failure analysis of belt conveyor damage caused by falling material. *Engineering Failure Analysis* , 36, 30-38.

[7] Wang Y., Miao C., Miao D., et al. (**2023**). Hazard source detection of longitudinal tearing of conveyor belt based on deep learning. *PLoS ONE* , 18(4), e0283878.

[8] Xiong N., Wen F., Li Q., et al. (**2022**). Review on detection technology of conveyor belt rip. *Coal Engineering* , 54(7), 79-85.

[9] Blazej R., Jurdziak L., Kirjanow A., & Kozlowski T. (**2015**). Evaluation of quality of steel cord belt splices. *Diagnostyka* , 16(3), 59-64.

[10] Peng X. (**2013**). *A novel image-based method for conveyor belt rip detection* . In 2013 IEEE International Conference on Signal Processing, Communication and Computing (ICSPCC 2013) (pp. 1-4). IEEE.

[11] Qiao T., Chen L., Pang Y., et al. (**2017**). Integrative binocular vision detection method based on infrared and visible light fusion. *Measurement*, 110, 192-201.

[12] Wang G., Zhang L., Sun H., & Zhu C. (**2021**). Longitudinal tear detection of conveyor belt under uneven light. *Measurement*, 168, 108341.

[13] Costa C., Antonucci F., Pallottino F., et al. (**2011**). Shape analysis of agricultural products: A review. Food and Bioprocess Technology, 4(5), 673-692.

[14] Yu L., Li Y., Luan Y., et al. (**2020**). Line structured light calibrating based on two-dimensional planar target. *Chinese Journal of Scientific Instrument* , 41(6), 124-131.

[15] Li X., Shen L., Ming Z., et al. (**2018**). Laser-based on-line machine vision detection for longitudinal rip. *Optik* , 168, 360-369.

[16] Xu H., Liu L., Shen K., et al. (**2021**). Longitudinal tear detection based on multi linear lasers. *Industry and Mine Automation*, 47(7), 37-44.

[17] Lv Z., Zhang X., Hu J., et al. (**2021**). Visual detection method based on line lasers. *Measurement*, 183, 109800.

[18] Zhao D., Zhang Z., Qin Y., et al. (**2025**). Research on longitudinal tear detection system based on line laser. *Coal Mine Machinery* , 46(1), 173-177.

[19] Shi Y., Shi H., Wang Y., et al. (**2024**). Research on longitudinal tear detection based on machine vision. *China Mining Magazine* , 33(4), 141-151.

[20] Zhang B. J., Hao H. T., Liu X. J., et al. (**2024**). An improved YOLOv5s-based foreign object detection algorithm for coal mine conveyor belts. Ningxia Engineering Technology, 23(2), 174-179.

[21] Du Z., Zhou H., Li C., et al. (**2022**). Small object detection based on deep CNNs: A review. *Computer Science* , 49(12), 205-218.

[22] Wang Y., Miao C., Miao D., et al. (**2023**). Hazard source detection based on deep learning. *PLoS ONE*,18(4), e0283878.

[23] Zhang M., Zhou M., Zhang Y., et al. (**2021**). Damage detection method based on deep learning. *Industry and Mine Automation* , 47(6), 51-56.

[24] Yu Q., Luo M., Xiang L., et al. (**2025**). Longitudinal tear detection based on line structured light. *Mining Safety & Environmental Protection* , 52(1), 180-186.

[25] He Y., Tian J., Zhang Z., et al. (**2023**). Lightweight research of YOLOv5 target detection. *Computer Engineering and Applications* , 59(1), 92-99.

[26] Hao S., Zhang X., Ma X., et al. (**2022**). Foreign object detection based on CBAM-YOLOv5. *Journal of China Coal Society* , 47(11), 4147-4156.

[27] Che J., Qiao T., Yang Y., et al. (**2021**). Longitudinal tear detection based on audio-visual fusion. *Measurement* , 176, 109152.

[28] Li W., Li C., & Yan F. (**2021**). Research on belt tear detection based on multiple laser line assistance. *Measurement* , 174, 109047.

[29] Zhou Y., Xu S., Huang Y., et al. (**2021**). Conveyor belt damage detection based on improved YOLOv4. *Industry and Mine Automation* , 47(11), 61-65.

[30] Hu J., Shen L., & Sun G. (**2018**). Squeeze-and-Excitation Networks. *Proceedings of the IEEE conference on computer vision and pattern recognition CVPR* , 7132-7141.

[31] You L., Zhu X., Chen Y., et al. (**2022**). Tear detection method of conveyor belt based on fully convolutional neural network. *Journal of Mine Automation* , 48(9), 16-24.