Computers and Artificial Intelligence；Vol.3, No.2; 2026
E-ISSN: 3006 -3728 P-ISSN：3006-371X
DOI: https://doi.org/10.70267/cai.26v3n2.7888
Published by: Zeus Press

BA-YOLO: A Lightweight Multi-scale Detection Model for Longitudinal Tear of Conveyor Belts

Jianguang Yin and Manli Wang*

Henan Polytechnic University, Jiaozuo, Henan, China

**Corresponding author: Manli Wang.*

Abstract

To enhance the multi-scale adaptability and real-time performance of the longitudinal tear detection network for conveyor belts, this paper proposes a lightweight tear detection algorithm based on laser line assistance, which is an improved version of the YOLOv11n model. Firstly, a bidirectional feature pyramid network (BiFPN) is introduced into the neck network to strengthen the network's ability to fuse features of different scales, enabling it to detect tear faults of different sizes on conveyor belts; Secondly, the downsample convolution layer of the main backbone network is replaced by an ADown layer, which reduces the size of the feature map and computational cost through a more efficient feature compression method, thereby effectively reducing the model complexity. Finally, the network is trained on the conveyor belt tear fault dataset. Experimental results show that the precision and recall rate of the BA-YOLO model are 92.4% and 82.8% respectively, and the mean average precision (mAP@0.5) is 90.3%. Compared with the original model, the model parameters have been reduced by 38.9% and the computational cost has been reduced by 14.3%. The improved model not only demonstrates the best detection performance but also shows stronger adaptability.

Keywords

conveyor belt tearing, multi-scale, yolov11, BiFPN, adown

1. Introduction

As a global energy pillar, coal production safety is vital to the stability of the energy industry. Belt conveyors are indispensable in mining transportation due to their long distance and continuous operation. However, under high-load underground conditions, conveyor belts are susceptible to longitudinal tears caused by foreign object friction, material aging, or heavy impacts. Undetected minor tears can escalate rapidly, leading to belt scrapping, significant economic losses, and severe threats to personnel safety. Therefore, achieving real-time and precise tear detection under complex backgrounds is a core requirement for intelligent mine construction [1].

Traditionally, tear monitoring has relied on manual inspection and contact-based physical devices such as mechanical baffles. Nevertheless, manual inspection is limited by human fatigue, while physical devices suffer from slow response times and poor interference resistance in harsh environments. Although deep learning-based computer vision offers a non-contact and high-precision alternative, mainstream models face

significant challenges in practical industrial applications. Stray light interference and dust occlusion hinder stable feature extraction, while the vast morphological variations of tears often lead to information loss during multi-scale feature fusion. Furthermore, excessive computational overhead limits execution efficiency in complex environments.

To address these challenges, this paper proposes a lightweight detection algorithm assisted by line laser, named BA-YOLO. Built upon the YOLOv11n baseline, the algorithm integrates the ADown downsampling module and the BiFPN feature fusion mechanism to reduce parameters while enhancing multi-scale distortion capture. The research hypotheses are as follows: First, line laser assistance improves contrast and suppresses feature blurring caused by stray light. Second, the ADown module reduces parameter redundancy while maintaining feature representation. Finally, the BiFPN mechanism facilitates cross-scale feature interaction, enhancing detection accuracy for narrow and complex tear morphologies.

2. Related Work

The technological evolution of conveyor belt tear detection reveals a progressive shift from rudimentary physical sensing to sophisticated computational frameworks. Early research primarily explored the deployment of discrete sensors and mechanical triggers. Kozlowski et al. pioneered a magnetic monitoring method for steel-cord joints, which evaluates belt integrity by analyzing magnetic responses [2]. Following this trajectory, the work of Dan et al. utilized ultrasonic conduction analysis, identifying internal damage by extracting signal variations during transmission through the belt structure [3]. Although these solutions established a baseline for industrial safety, they were frequently compromised by significant response latency and inherent vulnerability to corrosive underground environments. Subsequent innovations, despite their technical merits, remained plagued by high false-alarm rates. The persistent vibrations of heavy machinery and the accumulation of mineral dust often corrupted sensor readings, ultimately failing to satisfy the rigorous reliability standards required for modern automated operations.

The quest to eliminate the drawbacks of physical contact led researchers toward optical sensing and classical image processing. Li et al. advanced this field by utilizing an improved SSR algorithm to process images captured by line-scan cameras, effectively isolating tear features for real-time assessment [4]. While such techniques represented a pivotal step toward automation, their efficacy remained tethered to manually calibrated thresholds and idealized illumination. In the unpredictable conditions of an active mine, stray light and coal dust frequently obscure subtle geometric deformations. This loss of visibility results in unacceptably high rates of missed detections, demonstrating that traditional vision algorithms lack the adaptive robustness necessary to withstand the dynamic noise of industrial scenarios.

More recently, the rise of Convolutional Neural Networks (CNNs) has redefined defect detection through specialized architectural modifications. Single-stage detectors, particularly the YOLO family, have gained prominence for their ability to balance inference speed with end-to-end optimization. Liu et al. enhanced the perception of small-scale targets and regression stability by integrating the BotNet attention mechanism and Shape_IoU into the YOLOv5 framework. Parallel to these efforts [5], the methodology proposed by Wang et al. refined YOLOv7 through the introduction of the SimAM module and the SimSPPFCSPC model [6], utilizing the EIoU loss function to further elevate detection precision.

Despite the notable strides in these specialized studies, existing solutions still encounter two critical technical bottlenecks during field implementation. First, the challenge of extreme scale balancing persists; the vast cross-scale fluctuations of tear morphologies often cause subtle, narrow-width features to be overwhelmed by dominant background semantics. Second, a persistent conflict remains between detection accuracy and computational overhead. The tendency to stack complex attention modules significantly elevates inference latency, which in turn hinders the seamless deployment of algorithms on resource-constrained monitoring terminals. Consequently, there remains a distinct lack of a dedicated framework capable of concurrently managing stray light suppression, efficient cross-scale interaction, and model lightweighting, providing the primary impetus for the development of the BA-YOLO algorithm.

3. Research Design

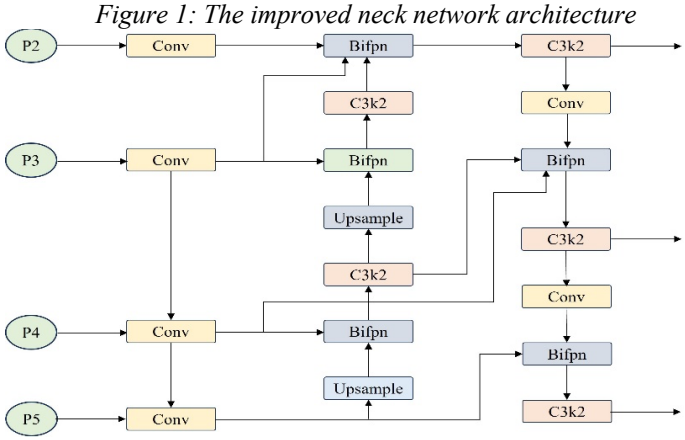
To address the inherent limitations of the baseline YOLOv11 [7] model when processing line-laser images of conveyor belts, this study introduces an optimized detection framework designated as BA-YOLO. The proposed architecture is specifically engineered to mitigate challenges such as extensive background redundancy, excessive computational latency, and the insufficient sensitivity of multi-scale feature fusion toward subtle tear signatures. While maintaining the high-speed inference characteristic of one-stage detectors, the BA-YOLO model undergoes a systematic reconstruction of its feature aggregation mechanism and spatial downsampling path to enhance its overall perception and localization capabilities.

The architectural refinement of BA-YOLO is implemented through two strategic modifications integrated into the neck and backbone of the network. In the feature integration stage, the BiFPN (Weighted Bidirectional Feature Pyramid Network) replaces the original PANet architecture within the neck. By introducing learnable weight parameters, BiFPN facilitates the adaptive fusion of features across disparate scales, ensuring that subtle tear characteristics are not overwhelmed by dominant global semantics. Subsequently, in the backbone stage, the conventional strided convolutions are replaced by the ADown efficient downsampling module. This component is designed to filter out redundant background regions while effectively preserving the high-frequency edge information of laser stripes. Through the synergy of these optimized modules, the network achieves a superior balance between detection precision and computational efficiency in complex industrial scenarios.

3.1 BiFPN Neck Network Architecture

Due to the varying shapes and sizes of the conveyor belt tears, different resolutions of features will be generated during the training process. The conventional linear superposition of these features may lead to uneven weight allocation of features related to different tear targets in the fused output. This imbalance may cause large-scale features to dominate the fused output, masking small-scale features, thereby potentially leading to missed detections. To solve this problem, it is necessary to consider preventing the loss of shallow features, as these features contain the key details for detecting small objects. Therefore, this study introduces a bidirectional feature pyramid network (BiFPN) module in the neck network, by constructing a bidirectional feature fusion mechanism, to enhance the model's ability to fuse tear detail features and global features. The improved neck network architecture is shown in Figure 1.

Figure 1: The improved neck network architecture



Compared with the traditional feature pyramid network, BiFPN has significantly optimized the topological connection structure. BiFPN adopts a bidirectional iterative topology of both top-down and bottom-up approaches, and through cross-scale skip connections and weighted feature fusion, it repeatedly aggregates information among multiple levels, effectively alleviating the problems of information loss and imbalance in deep networks. Without significantly increasing the model's parameter quantity, it achieves higher-order multi-scale feature reuse, effectively alleviating the feature degradation and detail loss problems during the forward propagation process of deep networks.

At the feature aggregation calculation level, BiFPN abandons the traditional equal fusion strategy and innovatively introduces a weighted feature fusion mechanism. Considering that the information gain of feature maps at different resolutions when representing specific scale targets is significantly different, the

network assigns scalar weights that can be independently learned to each fusion node. To balance the numerical stability of the training process and the hardware forward inference efficiency, this paper adopts a fast normalization fusion strategy instead of the traditional Softmax normalization. Its calculation formula is shown in Equation (1):

$$O = \sum_k \frac{\alpha_k I_k}{\varepsilon + \sum_l \alpha_l} \tag{1}$$

Take the aggregation process of the feature nodes in the 4th layer of the BiFPN structure as an example. It involves two independent feature interactions between the top-down intermediate feature P_4^{mid} and the bottom-up output feature P_4^{out}. Combined with the fast normalization strategy and the newly added cross-level connections within the same layer, the feature fusion process can be respectively represented by formulas (2) and (3):

$$P_4^{mid} = Conv\left(\frac{\alpha_1 \bullet P_4^{in} + \alpha_2 \bullet Resize(P_5^{in})}{\alpha_1 + \alpha_2 + \varepsilon}\right) \tag{2}$$

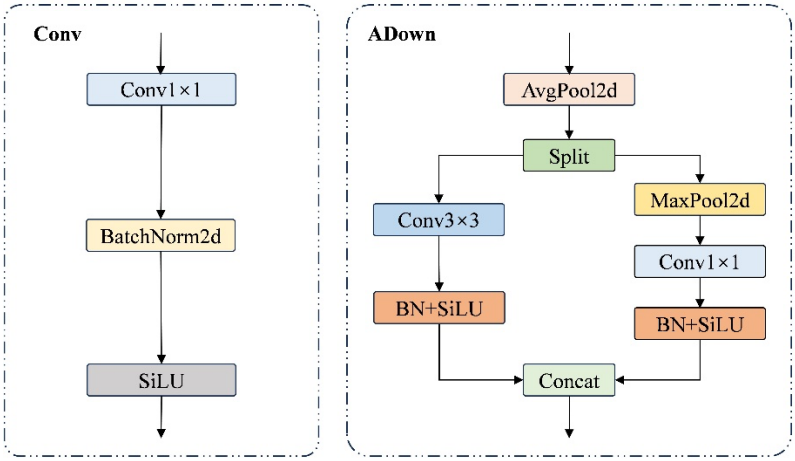
$$P_4^{out} = Conv\left(\frac{\alpha_1' \bullet P_4^{in} + \alpha_2' \bullet P_4^{mid} + \alpha_3' \bullet Resize(P_3^{out})}{\alpha_1' + \alpha_2' + \alpha_3' + \varepsilon}\right) \tag{3}$$

Among them, P_i^{in} represents the original input features transmitted from the main network to the i-th layer; Resize is the operation of upsampling or downsampling of the feature map performed to achieve spatial resolution matching; Conv represents the conventional convolution operation for feature processing; α_i and α_i' are independent learnable weight parameters for each branch.

3.2 Downsampling Module ADown

The images of conveyor belt tearing usually have high resolution and contain a large amount of background information. If the main network is directly used for feature extraction, it is prone to redundant computations and reduces the efficiency of the model. Therefore, this paper introduces the ADown module, which combines average pooling and maximum pooling operations, and can effectively reduce the computational complexity while retaining key information. By replacing the downsampling convolution structure in the traditional network with the ADown module, a lightweight design of the network is achieved, and the invalid computations are reduced while maintaining the integrity of the main semantic features. The comparison of the structures between the traditional convolution downsampling and the ADown module is shown in Figure 2.

Figure 2: Comparison of traditional convolution and ADown downsampling module structures



Unlike traditional methods that rely on a single convolutional path, the ADown module implements a dual-branch feature extraction and reorganization mechanism characterized by an initial smoothing stage followed by feature bifurcation. Let the input feature map of the current network layer be denoted as F_{in}. The ADown module first performs average pooling on F_{in} across the spatial dimensions. This operation is designed to extract a globally smoothed representation of the input features, thereby effectively suppressing high-frequency noise such as specular reflections from idlers and inherent conveyor belt textures. The mathematical formulation is as follows:

$$F_{avg} = AvgPool2d(F_{in}) \tag{4}$$

Subsequently, the downsampled smooth feature map F_{avg} is fed into two parallel processing branches. The main branch utilizes 3×3 convolutions to extract deep local structural information, followed by a sequential arrangement of Batch Normalization (BN) layers and SiLU activation functions. This design not only strengthens the network's nonlinear mapping capability regarding the geometric distortion of laser stripes but also enhances the numerical stability of the model during the training process.

$$G_{avg} = Conv_{3\times3}(F_{avg}) \tag{5}$$

The auxiliary branch incorporates a parallel 3×3 max-pooling layer, leveraging its high sensitivity to local extrema to specifically capture salient feature responses from tear edges and high-intensity laser regions. Subsequently,1×1 convolutions are employed to achieve channel-wise compression and cross-channel information interaction, thereby obtaining a more compact feature representation:

$$F_{max} = MaxPool2d(F_{avg}) \tag{6}$$

$$G_{\max} = Conv_{1\times1}(F_{\max}) \tag{7}$$

Finally, the output features from the main and auxiliary branches are fused via concatenation (Concat) along the channel dimension to generate a multi-scale output feature F_2 with rich representational capacity. This fused feature not only encompasses the global contextual semantics dominated by average pooling but also integrates the local edge details captured by max pooling. The computation process is formulated as shown in equation (8):

$$F_{out} = Concat(G_{avg}, G_{\max}) \tag{8}$$

In summary, through the aforementioned parallel aggregation and decoupled computation mechanism, the ADown module achieves a reduction in spatial resolution while significantly cutting down redundant convolutional operations in non-informative background regions. Furthermore, by virtue of a complementary feature extraction strategy that combines mean-smoothing with extrema-preservation, the module maximizes the retention of high-frequency gradient information for tear features. This provides a superior data foundation for the subsequent construction of the multi-scale feature pyramid.

4. Empirical Analysis

To verify the effectiveness of the proposed improved algorithm, two sets of experiments were conducted:

In the ablation study, the original YOLOv11n model was first employed as the baseline for benchmarking. Subsequently, the BiFPN neck architecture and the ADown downsampling module were integrated into the baseline both independently and jointly to evaluate their respective contributions to model performance. This approach allows for a quantitative analysis of each module's impact on detection accuracy and computational efficiency.

In the comparative experiments, the performance of the improved YOLOv11n was evaluated against various state-of-the-art (SOTA) object detection models to demonstrate its superiority under identical task conditions. By comparing with these mainstream benchmarks, the advantages of the proposed algorithm in terms of detection precision, inference speed, and computational efficiency are clearly illustrated.

4.1 Experimental Environment and Hyperparameter Settings

All network training and testing in this study were conducted on a unified experimental platform. The operating system employed was CentOS Linux 7.6.1810, equipped with an NVIDIA GeForce RTX 3090 GPU (24 GB VRAM). The software environment consisted of the PyTorch 2.0.0 deep learning framework, Python 3.9.20 interpreter, and CUDA 11.8 toolkit. To verify the feasibility and superiority of the proposed improved network model, a specialized conveyor belt longitudinal tear simulation apparatus was constructed, as illustrated in Figure 3.

Figure 3: Simulation apparatus for conveyor belt tear failures

To ensure the comparability of the experimental results, a unified hyperparameter configuration was adopted for all models in this study, with the specific settings detailed in Table 1. These hyperparameters were kept consistent throughout the training process to ensure that the performance of different models was evaluated under identical conditions.

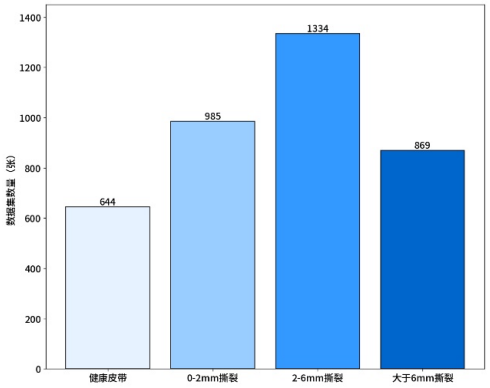
Table 1: Hyperparameter settings for model training

Parameter	Value
Epochs	200
Initial learning rate	0.01
Batch size	16
Optimizer	SGD
Weight decay	0.0005
Momentum	0.937

4.2 Dataset

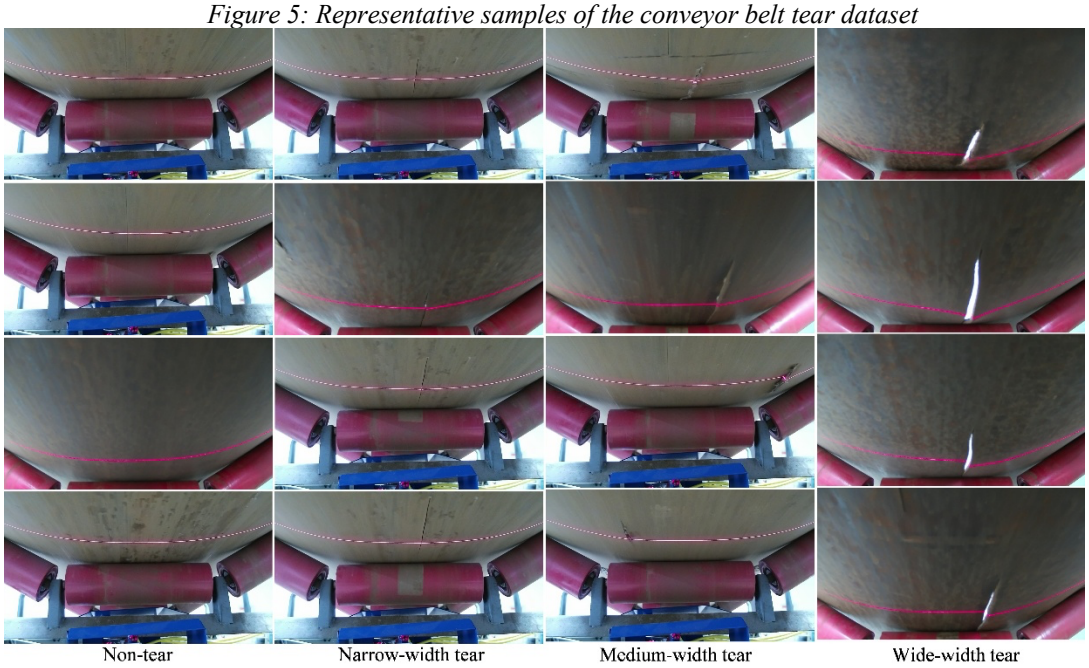
The data used in the experiments were collected from the laboratory-constructed fault simulation apparatus, where synchronous dynamic capture was performed on the tear regions of a running conveyor belt. Images were extracted frame-by-frame from the captured videos, followed by a filtering process to remove redundant or substandard samples. After data augmentation, the final dataset comprised 3,832 conveyor belt images, including both healthy samples without defects and those featuring tear failures. To evaluate the effectiveness of the improved algorithm across various tear scales, the tear images were categorized into three types based on crack width: narrow-width tears (0–2 mm), medium-width tears (2–6 mm), and wide-width tears (exceeding 6 mm). The distribution of the dataset is illustrated in Figure 4.

Figure 4: Distribution of the dataset across different tear categories



The dataset was randomly partitioned into training, validation, and testing sets with a ratio of 7:2:1. The annotation for the training and validation sets was performed using the LabelImg tool. Samples from the dataset are illustrated in Figure 5.

Figure 5: Representative samples of the conveyor belt tear dataset



4.3 Ablation Study and Analysis

To evaluate the impact of each proposed module on model performance, we conducted a series of ablation studies based on the YOLOv11n baseline, incorporating the BiFPN neck architecture and the ADown module for comparative analysis. A comprehensive set of evaluation metrics was employed to quantify the model's performance, including Parameters, GFLOPs (Giga Floating-Point Operations), Precision, Recall, mAP@0.5 (Mean Average Precision at an IoU threshold of 0.5), and mAP@0.5:0.95 (Mean Average Precision averaged over IoU thresholds ranging from 0.5 to 0.95).

The experiments were conducted using the baseline YOLOv11n network, along with the integration of BiFPN and ADown modules. The ablation study results are summarized in Table 2, where "✓" denotes the inclusion of the corresponding method.

Table 2: Results of the ablation study on the conveyor belt tear dataset

Model			Parameters/M	GFLOPS	Precision/%	Recall/%	mAP@0.5/%	mAP@0.5:.95/%
Baseline	BiFPN	ADown						
✓			2.582	6.3	89.2	77.3	82.1	35.5
✓	✓		1.923	6.3	89.1	83.2	86.7	35.8
✓		✓	2.103	5.3	90.0	82.2	85.0	39.3
✓	✓	✓	1.578	5.4	92.4	82.8	90.3	38.4

Table 2 illustrates the performance evolution after integrating the BiFPN and ADown modules into the YOLOv11n baseline. Overall, the proposed structural refinements enhance detection efficacy to varying degrees while simultaneously reducing model parameters and computational costs.

The experimental results demonstrate that the introduction of the BiFPN module achieves structural optimization while maintaining high detection performance. By facilitating efficient information transmission across multi-scale feature maps, BiFPN bolsters the model's capability to perceive targets of various sizes. Compared with the baseline, this configuration increases Recall and mAP@0.5 by 5.9% and 4.6%, respectively, while reducing the parameter count by 25.5% (0.659 M). This improvement is attributed to the bidirectional multi-scale fusion architecture, which strengthens the interaction between shallow-level details and deep-level semantics. Although a marginal decrease in Precision is observed—likely due to the

inclusion of more boundary-ambiguous samples following the Recall enhancement—the overall performance maintains a distinct upward trend.

Furthermore, the integration of the ADown module yields even more pronounced lightweighting benefits. By employing a dual-branch structure of average and max pooling, ADown effectively preserves salient local features during spatial downsampling, ensuring stable feature representation. With this module, the parameters and GFLOPs are reduced by 18.6% (0.479 M) and 15.9%, respectively, while mAP@0.5:0.95 increases by 4.4%. These results indicate that ADown successfully reduces computational overhead while enhancing the quality of backbone features and strengthening the response to weak features and minute targets.

Ultimately, the simultaneous integration of both BiFPN and ADown enables the model to reach the optimal balance between efficiency and accuracy. The model achieves a 1.007 M (39.0%) reduction in parameters while boosting mAP@0.5 by 8.2% to reach 90.3%. This success stems from the structural complementarity between the two modules: ADown preserves the integrity and stability of downsampled features in the backbone, while BiFPN reinforces cross-scale information interaction in the neck. Together, they establish a synergistic mechanism between high-quality feature representation and efficient multi-scale fusion, allowing the proposed model to achieve peak detection performance with substantially lower computational requirements. In summary, the ablation study fully validates that each proposed improvement contributes positively to the various evaluation metrics for conveyor belt tear detection.

4.4 Comparative Experiments and Analysis

To comprehensively and objectively evaluate the performance of the proposed BA-YOLO model in practical conveyor belt longitudinal tear detection, this section conducts comparative experiments with several representative object detection algorithms. The selected baseline models encompass a variety of network architectures of different scales, specifically including YOLOv5n, YOLOv8n, YOLOv9t, YOLOv9s [8], YOLOv10n, YOLOv10s [9], and the original YOLOv11n. To ensure a fair and rigorous evaluation, all models were trained and validated under identical hardware environments, hyperparameter configurations, and the same conveyor belt longitudinal tear dataset. The specific performance evaluation results for each algorithm are summarized in Table 3.

Table 3: Results of the comparative experiments on the conveyor belt tear dataset

Model	Parameters/M	GFLOPS	Precision/%	Recall/%	mAP@0.5/%	mAP@0.5:.95/%
YOLOV5	2.182	5.8	0.867	0.792	0.824	0.352
YOLOV8n	2.684	6.8	0.861	0.813	0.842	0.368
YOLOV9t	1.730	6.4	0.833	0.785	0.831	0.347
YOLOV9s	6.194	22.1	0.899	0.814	0.870	0.410
YOLOV10n	2.695	8.2	0.831	0.735	0.792	0.337
YOLOV10s	8.036	24.4	0.873	0.738	0.834	0.358
YOLOV11n	2.582	6.3	0.893	0.804	0.852	0.370
BA-YOLO	1.577	5.4	0.924	0.828	0.903	0.384

In terms of quantitative complexity metrics, BA-YOLO demonstrates a significant lightweight profile. With a parameter count of 1.577 M and GFLOPs of 5.4, the proposed model maintains the lowest computational requirements among all evaluated baselines. Compared to the original YOLOv11n architecture, BA-YOLO achieves a reduction in parameter scale by approximately 38.9%, with a corresponding decrease in computational overhead. These results underscore the efficacy of the ADown parallel pooling module in eliminating redundant background computations and streamlining the network topology. Such efficiency provides a robust foundation for the deployment of the algorithm on resource-constrained edge devices within underground mining environments.

Regarding the quantitative evaluation of detection accuracy, the comprehensive performance of mainstream models is often constrained by their general-purpose design. These architectures typically lack targeted optimization for the localized distortions and multi-scale tear features characteristic of conveyor belts. As shown in Table 3, the mAP@0.5 for most baseline models ranges between 79% and 87%. In contrast, the BA-YOLO model yields a Precision of 92.4% and a Recall of 82.8%, resulting in a superior mAP@0.5 of 90.3%. Notably, on the more stringent mAP@0.5:0.95 metric—which measures the tightness

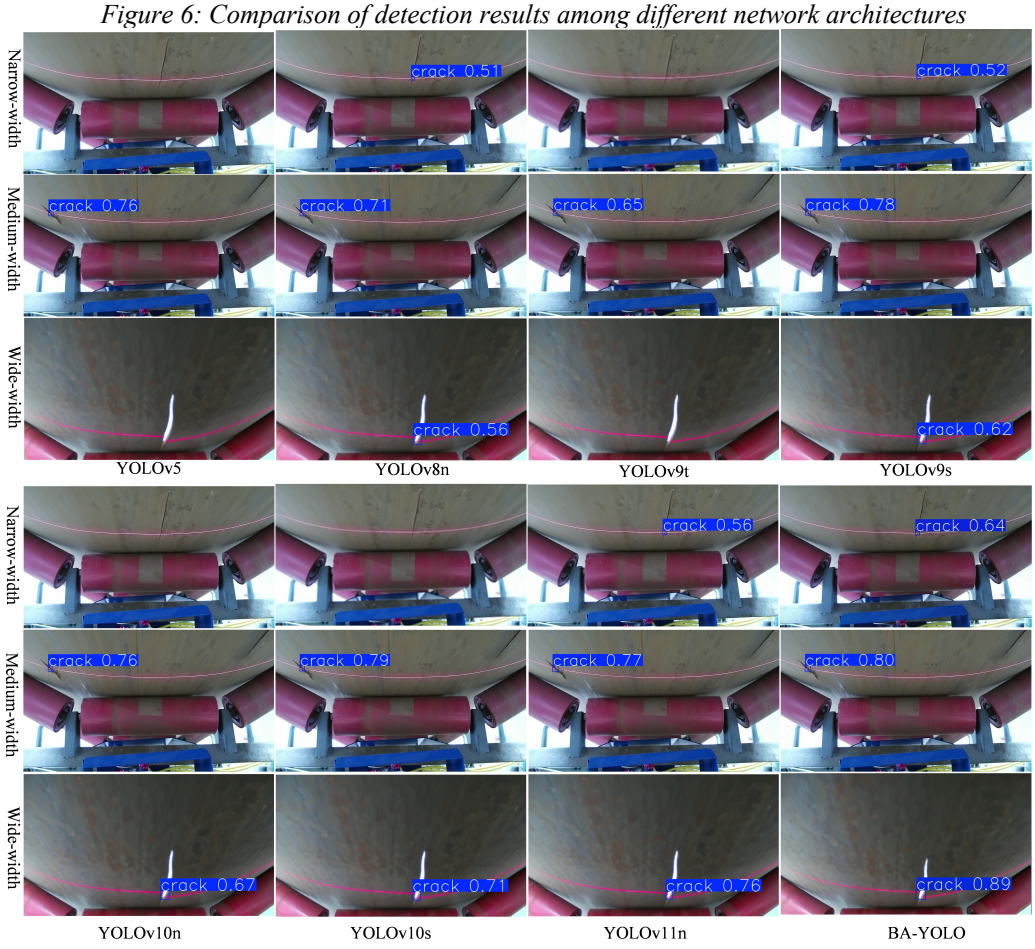
of bounding box regression—the performance of BA-YOLO stands at 38.4%. This value surpasses that of the baseline YOLOv11n and most other lightweight frameworks, further validating the robustness of the proposed strategy in complex industrial detection tasks.

In terms of quantitative complexity metrics, BA-YOLO demonstrates a significant lightweight profile. The parameter count of the proposed model is 1.577 M, and the GFLOPs are 5.4, both of which represent the lowest levels among all evaluated baselines. Compared to the original YOLOv11n architecture, BA-YOLO achieves a reduction in parameter scale by approximately 38.9%, with a corresponding decrease in computational overhead. These results underscore the efficacy of the ADown parallel pooling module in eliminating redundant background computations and streamlining the network topology, validating the feasibility of high-efficiency execution on resource-constrained edge devices.

Regarding the quantitative evaluation of detection accuracy, the comprehensive performance of mainstream models is often constrained by their general-purpose design. These architectures typically lack targeted optimization for localized distortions and multi-scale features inherent in the specific dataset. As shown in Table 3, the mAP@0.5 for most baseline models fluctuates between 79% and 87%. In contrast, the BA-YOLO model yields a Precision of 92.4% and a Recall of 82.8%, resulting in a superior mAP@0.5 of 90.3%. Notably, for the more stringent mAP@0.5:0.95 metric, which measures the tightness of bounding box regression, the performance of BA-YOLO stands at 38.4%. This value surpasses that of the baseline YOLOv11n and most other lightweight frameworks, further validating the robustness of the proposed strategy in complex detection tasks.

To verify the generalization capability and interference resistance of each algorithm in complex physical scenarios, representative samples featuring narrow, medium, and wide tears were extracted from the test set. A qualitative visual analysis was then conducted on the prediction results of the aforementioned baseline models, with the specific detection performance illustrated in Figure 6.

Figure 6: Comparison of detection results among different network architectures



As illustrated in Figure 7, detecting narrow tears is extremely challenging due to the faint features and significant light scattering. Under these conditions, YOLOv5, YOLOv9t, YOLOv10n, and YOLOv10s fail to effectively extract the high-frequency features of local fractures, resulting in severe missed detections. Although YOLOv8n, YOLOv9s, and YOLOv11n identify the targets, their classification confidence scores remain low, ranging from 0.51 to 0.56. In contrast, BA-YOLO accurately localizes the distorted regions and achieves the highest confidence score (0.64) among the tested models.

For medium and wide tears with more distinct features, most models achieve basic localization; however, significant disparities exist in their confidence scores. Wide-width scenarios, in particular, often trigger scale-adaptability failures, leading to repeated missed detections by YOLOv5 and YOLOv9t, which underscores their limitations in handling drastic cross-scale features. Conversely, the prediction results of BA-YOLO are not only highly consistent with the laser fracture edges but also yield high confidence scores of 0.80 and 0.89 in medium and wide-width scenarios, respectively. These visualization results intuitively demonstrate the superior performance of BA-YOLO in multi-scale object detection and its robust interference resistance in complex environments.

5. Conclusion

To ensure operational safety and address the limitations of existing detection methods under complex industrial conditions, such as stray light interference, dust occlusion, and drastic cross-scale feature variations, this paper proposes a lightweight detection algorithm assisted by line laser, named BA-YOLO. Built upon the YOLOv11n baseline, the network architecture incorporates two pivotal enhancements. First, the lightweight downsampling module ADown is introduced into the backbone to minimize computational overhead while preserving essential feature information through a parallel pooling strategy. Second, a BiFPN-based neck is employed to reconstruct the feature fusion network, which facilitates efficient cross-scale interaction and weighted integration of multi-level features. This enhancement significantly bolsters the model's capability to detect targets across varying scales and complex morphologies. Experimental results demonstrate that the proposed BA-YOLO model achieves a high detection accuracy, with an mAP@0.5 of 90.3%, while simultaneously reducing the parameter count by approximately 38.9% compared to the baseline. These findings validate the superior balance between computational efficiency and detection performance, offering a robust solution for high-performance real-time monitoring.

References

[1] You L, Luo M, Zhu X, et al. Deep encoder-decoder networks for belt longitudinal tear detection[J]. Measurement and Control, 2025, 58(5): 643-655.

[2] Kozłowski T, Błażej R, Jurdziak L, et al. Magnetic methods in monitoring changes of the technical condition of splices in steel cord conveyor belts[J]. Engineering Failure Analysis, 2019, 104: 462-470.

[3] Dobrotă D. Vulcanization of rubber conveyor belts with metallic insertion using ultrasounds[J]. Procedia Engineering, 2015, 100: 1160-1166.

[4] Li J, Miao C. The conveyor belt longitudinal tear on-line detection based on improved SSR algorithm[J]. Optik, 2016, 127(19): 8002-8010.

[5] Liu W, Tao Q, Wang N, et al. YOLO-STOD: an industrial conveyor belt tear detection model based on Yolov5 algorithm[J]. Scientific Reports, 2025, 15(1): 1659.

[6] Wang Y, Du Y, Miao C, et al. Longitudinal tear detection of conveyor belt based on improved YOLOv7[J]. IEEE Access, 2024, 12: 24453-24464.

[7] Khanam R, Hussain M. Yolov11: An overview of the key architectural enhancements (2024)[J]. arXiv preprint arXiv:2410.17725, 2024.

[8] Wang C Y, Yeh I H, Mark Liao H Y. Yolov9: Learning what you want to learn using programmable gradient information[C]//European conference on computer vision. Cham: Springer Nature Switzerland, 2024: 1-21.

[9] Wang A, Chen H, Liu L, et al. Yolov10: Real-time end-to-end object detection[J]. Advances in neural information processing systems, 2024, 37: 107984-108011.

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgment

This paper is an output of the science project.