Sensors and Materials, Vol. 38, No. 9 (2026) 5071–5090
MYU Tokyo

S & M 4611

Visual Selection of Shredded Tobacco with Improved You-Only-Look-Once-Version-11-Nano and Its Width Measurement

Yuxin Dai,[1] Feng Guo,[1] Guodong Lin,[1] Chen Cai,[2]
Yuhang Hu,[2] Zhiqiang Chen,[2] and Haibin Wu[2*]

[1]Xiamen Tobacco Industrial Co., Ltd., Xiamen 361022, China
[2]School of Mechanical Engineering and Automation Fuzhou University 350116, China

(Received March 23, 2026; accepted June 29, 2026)

Keywords: shredded tobacco, visual selection, YOLO11n, BiFPN, dynamic head, scale-based loss, width measurement

The width of shredded tobacco is an important quality indicator in cigarette manufacturing, but reliable measurement is difficult because tobacco shreds are often fine, curled, fragmented, and densely overlapped. In this study, we propose a You-Only-Look-Once-Version-11-Nano (YOLO11n)-based visual selection and width measurement framework for identifying isolated and morphologically suitable tobacco shreds from complex production-line images. The proposed model integrates a bidirectional feature pyramid network, dynamic head, and scale-aware bounding box regression loss to improve multi-scale feature fusion, feature representation, and localization stability for slender targets. The images collected from an actual tobacco primary processing line were annotated according to measurement suitability and used for model evaluation. The improved model achieved a mean average precision (*mAP*) of 86.61% and an *F*1-*score* of 0.7935, representing improvements over the original YOLO11n of 11.15 percentage points in *mAP* and 0.0794 in *F*1-*score*, with an inference speed of 19.19 frames per second. The selected tobacco shred regions were further measured by a variable-radius circle method. Field validation on three production batches, covering 300 tobacco shreds, showed an average relative error of 4.66%, below the enterprise acceptance threshold of 8%. These results demonstrate that the proposed framework can effectively support tobacco shred width inspection under tested production-line conditions.

1. Introduction

Tobacco cutting is a key step in cigarette production. The cutting quality, particularly the tobacco shred width and its variance, is of significant importance for quality control and process stability. Tobacco shred width not only affects ignition uniformity, flammability, and smoke flow resistance but also directly relates to taste consistency.(1) Therefore, the precise

*Corresponding author: e-mail: wuhb@fzu.edu.cn
https://doi.org/10.18494/SAM6347

ISSN 0914-4935
https://myukk.org/

measurement of shred width, the extraction of variance characteristics, and the early warning of abnormal widths are critical in the quality control of actual production.(2)

Early detection methods based on traditional image processing often relied on manual sampling combined with optical microscopic measurements or on algorithms using geometric-feature-based rules to estimate tobacco shred width. Zhao *et al.*(3) developed a tobacco shred width measurement device based on a linear-array CCD, achieving a relative error below 0.5% and requiring only one-sixth of the time needed by a projector-based measurement method. Liu *et al.*(4) proposed a detection scheme based on variable-diameter circles. This scheme determined the shred width by measuring the vertical distance between the diameter of an inscribed circle and the centerline of a shred sample, then averaging the diameters of inscribed circles at multiple positions to obtain the final width. Such methods can achieve relatively intuitive quantitative analysis under conditions with simple backgrounds, single-layer shred arrangement, and stable lighting. However, they suffer from low sampling efficiency, typically require measuring shreds one by one, have slow measurement speeds, and are generally only suitable for offline periodic sampling inspection.

With the continuous development of tobacco-cutting technology, the requirements for the accuracy and efficiency of cut-width detection are increasing. The online detection of cutting width is bound to become an essential guarantee for improving the real-time performance of quality monitoring and feedback, and for achieving informationalized and digitalized production. However, online width measurement in dense arrangements remains challenging because tobacco shreds sampled from the production line are often entangled, occluded, or overlapped, making individual shreds difficult to separate and measure reliably. Before width measurement, it is crucial to select suitable tobacco shreds as detection objects.

In recent years, deep learning algorithms have achieved revolutionary advances in computer vision, enabling new solutions for visual detection in complex industrial scenarios. Object detection algorithms such as Faster Region-based Convolutional Neural Network (Faster R-CNN), Single Shot Detector, and You Only Look Once (YOLO) series have been widely applied in industrial defect detection and precision dimensional measurement tasks owing to their powerful feature extraction and end-to-end recognition and localization capabilities. Compared with traditional methods, deep learning models can automatically learn deep features from tobacco shred images and demonstrate greater robustness against complex conditions such as lighting variations, shred overlap, and background interference. Hu *et al.*(5) used a Caffe model to identify tobacco stems and shreds, then encoded the edge pixel points of the segmented and contour-extracted shred images. Inscribed circles were drawn, and their diameters were used as shred widths. Liu *et al.*(6) proposed a neural network algorithm with multiple convolutional layers for target recognition and width acquisition for tobacco shreds, demonstrating high contour recognition performance and width detection accuracy. However, these methods have relatively strict environmental requirements for shred measurement, making integration into actual production lines difficult. Zhao and Hu(7) realized the real-time detection of tobacco shred width using an improved YOLOv4 algorithm and the moving center search method, but the detection accuracy was relatively low.

In this paper, we propose a method for selecting measurement-suitable tobacco shreds from scattered distributions, followed by the width measurement of the selected shreds. We introduce three modifications to YOLO11n for the selection of tobacco shreds. The main improvements include the following: (1) adding a Bidirectional Feature Pyramid Network (BiFPN) in the neck network to achieve the effective fusion of multi-scale features[8] and enhancing the feature representation capability for slender targets; (2) adopting a dynamic head in the detection head, which enhances the perception of shred end features through three attention mechanisms: scale-aware, spatial-aware, and task-aware;[9] and (3) replacing the traditional CIoU Loss with a Scale-based Dynamic Loss for the Bounding Box (SDB Loss), which introduces a scale-aware dynamic weight allocation mechanism to specifically enhance the model's detection capability for small targets.[10] After the selection of the tobacco shreds, the width can be measured on the basis of a variable-radius circle algorithm. Experimental results showed that this method improves the accuracy and robustness of tobacco shred width measurement, meeting the efficiency and precision requirements for quality inspection in the tobacco online production.

2. Methodology

Prior to detailing the deep learning architecture for tobacco shred selection, it is necessary to establish the criteria for visual screening. In the primary tobacco processing line, variations in raw materials and cutting processes induce extreme morphological diversity among the shreds. The inherent tendency of these shreds to agglomerate and entangle poses a formidable challenge for automated online sampling and width measurement. To address this, an automated multi-stage sampling system is deployed (Fig. 1). A collaborative robotic arm equipped with a suction cup acquires shred samples from the conveyor belt [Fig. 1(a)], with the acquisition interface detailed in the close-up view [Fig. 1(b)]. The extracted sample is then deposited onto a vibrating sieve for size-based pre-filtering, after which the appropriately sized shreds slide through a chute onto a vibrating plate [Fig. 1(c)]. However, even after undergoing this rigorous physical dispersion process, the resulting visual distribution [Fig. 1(d)] demonstrates that only a marginal proportion of the shreds can be effectively isolated. Furthermore, the retrieved samples inevitably contain substantial particulate debris and excessively short fragments. These morphologically irregular instances fail to represent the authentic cutting width, rendering them invalid for measurement. Consequently, the robust visual selection of morphologically intact and

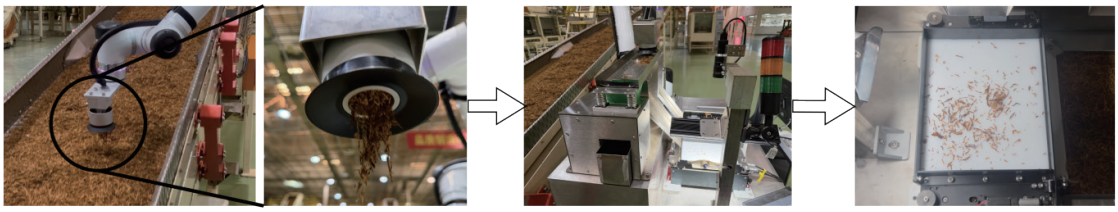

Fig. 1. (Color online) Automated online sampling and physical dispersion workflow for tobacco shreds. (a) Robotic arm acquiring samples directly from the production conveyor belt. (b) Close-up view of the suction cup capturing shred samples. (c) Shreds being prefiltered by a vibrating sieve and sliding onto the vibrating plate. (d) Final distribution of shreds after mechanical vibration dispersion.

isolated shred samples is a critical prerequisite for reliable width evaluation and quality control in tobacco manufacturing.

Therefore, the established criteria for visual selection prioritize shreds that are physically isolated, morphologically uniform, and slender, with minimal curling or folding. Figure 2 illustrates the distinct morphological contrast between shred samples suitable for width measurement, as indicated by the blue arrows, and those that are unqualified. The primary objective of defining these visual selection criteria is to accurately filter and locate these high-quality candidate targets amidst complex backgrounds prior to the width measurement phase. To fulfill this objective, an improved YOLO11n object detection framework is designed.

2.1 Improvement framework of YOLO11n

A method is proposed for selecting isolated samples from clustered tobacco shreds based on an improved YOLO11n architecture. As the lightweight version within the YOLO11 series, the baseline YOLO11n is a single-stage object detection framework that accomplishes target localization and category prediction within an end-to-end network. Its architecture consists of three main components:[11] the Backbone, which efficiently extracts multi-level features from the input image; the Neck, which fuses features from different scales; and the Head, which generates bounding box coordinates and class probabilities.

However, given the distinctive morphology of tobacco shreds, specifically their slender shape, dense overlap, and the high demand for boundary localization accuracy, the standard YOLO11n struggles to maintain precision in complex online scenarios. To address these challenges, we introduce three key modifications to the baseline architecture: (1) integrating BiFPN in the neck to enhance multi-scale feature fusion; (2) adopting a Dynamic Head with attention mechanisms to capture detailed endpoint features; and (3) employing SDB Loss to optimize regression accuracy for small targets. The structural comparison between the original YOLO11n and the improved network is illustrated in Fig. 3. The left side of the figure depicts the standard architecture using traditional Concat modules and CIoU loss, whereas the right side highlights the proposed improvements.

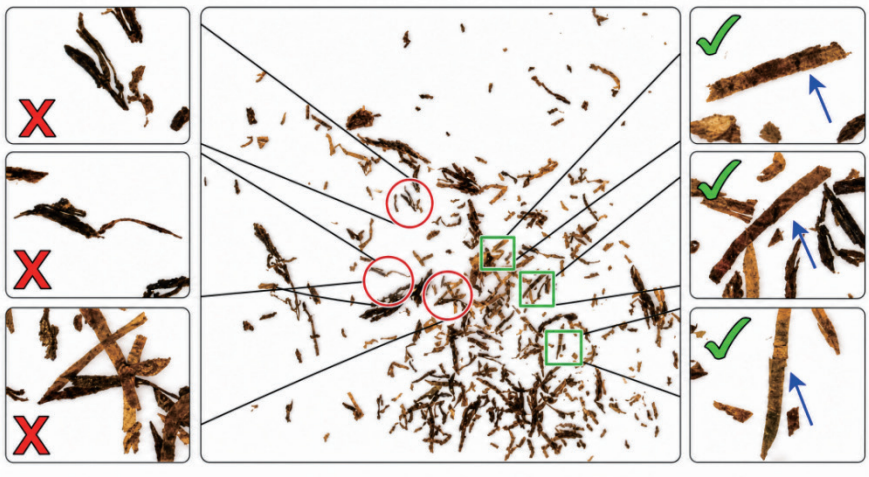

Fig. 2. (Color online) Morphological comparison between suitable (right, green boxes) and unsuitable (left, red circles) tobacco shreds.

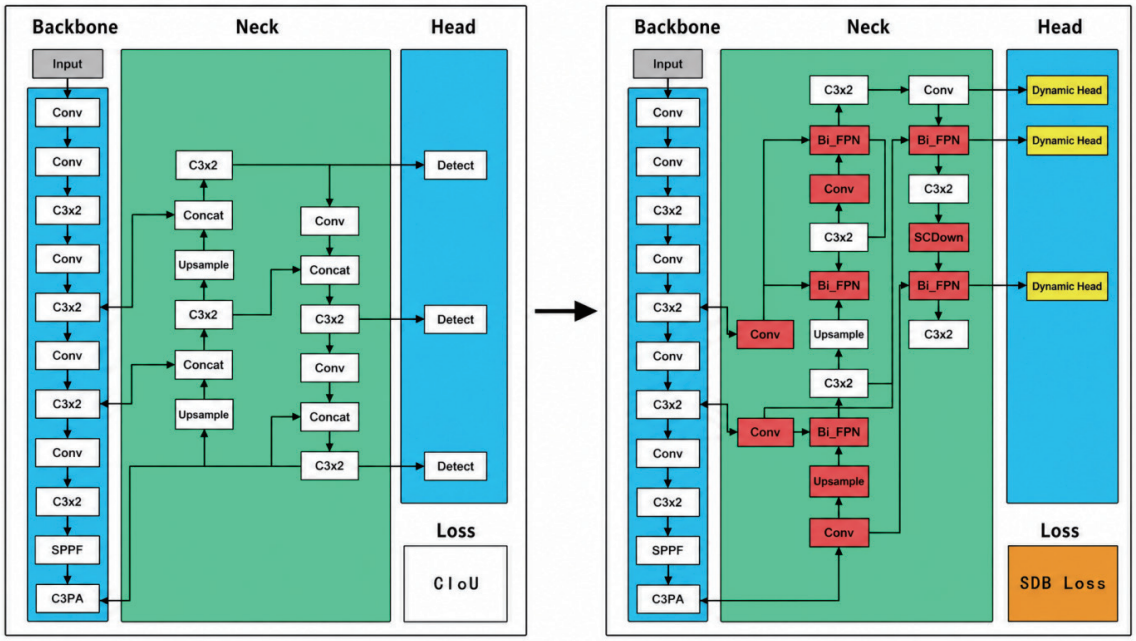

Fig. 3. (Color online) Comparison of the original and improved YOLO11n network architectures.

2.2 BiFPN

In the Neck network, the original Concat module is replaced with a BiFPN module. The original Concat operation merely stacks feature maps of different scales along the channel dimension, implicitly assuming equal importance for all input features. However, for small targets such as tobacco shreds, high-resolution shallow features, rich in detailed and positional information, are far more critical than deep features containing abstract semantic information. BiFPN addresses this by introducing learnable weighting parameters, enabling the network to adaptively learn and assign varying importance weights to features at different resolutions during training. This mechanism allows the model to focus more on high-detail features crucial for tobacco shred detection, thereby significantly enhancing the feature extraction and retention capability for minute shred segments.

BiFPN is an efficient multi-scale feature fusion architecture derived from the traditional Feature Pyramid Network (FPN).[12] Its core design principle is to substantially improve multi-scale feature fusion performance by establishing bidirectional cross-scale connections and incorporating a learnable weighting mechanism. Unlike the traditional FPN that employs only a top-down unidirectional pathway, BiFPN facilitates the entire interaction between high-level semantic information and low-level detailed features through bidirectional information flow.

BiFPN incorporates three key structural optimizations: First, nodes with only a single input edge are removed, simplifying the network structure and reducing computational redundancy. Second, shortcut connections are added between input and output nodes at the same level, promoting feature reuse and gradient flow.[13] Finally, each bidirectional path is designed as a repeatable, stackable feature network layer; stacking these layers multiple times enhances the model's multi-scale modeling capability. The structure of BiFPN is illustrated in Fig. 4.

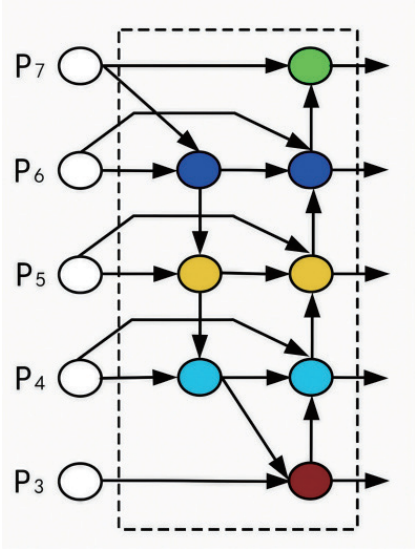

Fig. 4. (Color online) BiFPN structure diagram.

BiFPN introduces a learnable weighted feature-fusion mechanism to adaptively balance the contributions of input feature maps from different resolution levels. Let P_{in}^{i} denote the i-th input feature map, where i = 1, 2, ..., n, and n is the total number of input feature maps involved in one fusion node. For each input feature map, a learnable scalar weight w_i is assigned to represent its relative importance. To avoid negative fusion weights and improve numerical stability, the weights are normalized using a fast normalized fusion strategy:

$$\hat{w}_i = \frac{w_i}{\varepsilon + \sum_{j=1}^{n} w_j}, \tag{1}$$

where $w_i \geq 0$ is the learnable weight of the i-th input feature map, $\hat{w}_i$ is the normalized fusion weight, j is the summation index, and ε is a small positive constant used to prevent division by zero.

The weighted fused output feature map P_{out} is then computed as the weighted summation of all input feature maps:

$$P_{out} = \sum_{i=1}^{n} \hat{w}_i P_{in}^{i}, \tag{2}$$

where P_{out} denotes the output feature map after BiFPN fusion. Through this normalization mechanism, BiFPN enables the network to learn the relative importance of multi-scale features and enhances the representation of slender tobacco shreds at different spatial resolutions.

2.3 Dynamic Head

In the Head network, the standard Detect head is replaced with a Dynamic Head module, which offers the advantage of introducing dynamic perception and adaptive computation capabilities. Tobacco shreds do not exhibit standard rectangular shapes but often appear as slender, curved, or irregular fragments. The fixed parameters of traditional detection heads struggle to fit the boundaries of such targets accurately.[14] Dynamic Head incorporates three attention mechanisms: scale-aware, spatial-aware, and task-aware. This enables dynamic focus on key regions and detailed features across different dimensions, effectively enhancing recognition capability for small targets with irregular shapes or under partial occlusion. During the actual detection of tobacco shred width, phenomena such as shred stacking and edge curling can easily lead to false detections. By leveraging a more refined feature representation, Dynamic Head significantly suppresses background and noise interference, thereby reducing the false-positive rate. The structure of Dynamic Head is illustrated in Fig. 5.

In the Dynamic Head structure, three attention modules are cascaded sequentially to enhance feature representation from the level, spatial, and task dimensions. Specifically, the scale-aware module focuses on feature responses across different feature pyramid levels, the spatial-aware module adaptively highlights informative spatial locations, and the task-aware module dynamically adjusts channel responses for classification and localization tasks. Given an input feature tensor F, the overall attention mapping of Dynamic Head can be expressed as

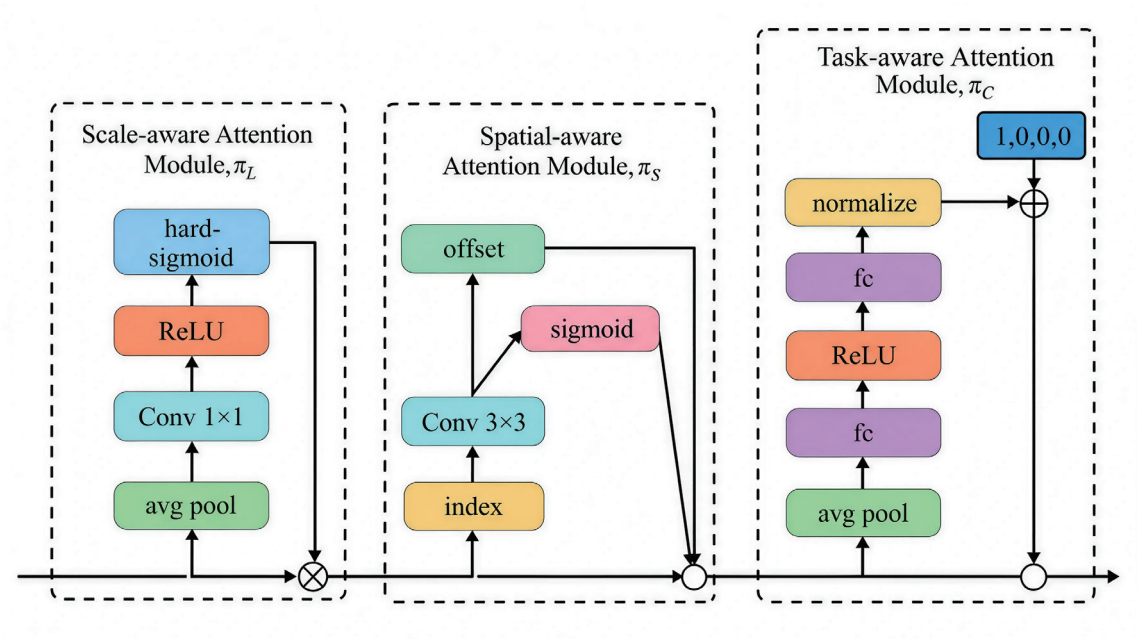

Fig. 5. (Color online) Dynamic Head structure diagram.

$$W(F) = \pi_C\left(\pi_S\left(\pi_L(F)\cdot F\right)\cdot F\right)\cdot F\,, \tag{3}$$

where $W(F)$ denotes the output feature after dynamic attention enhancement, F is the input feature tensor, and π_L, π_S, and π_C represent the scale-aware, spatial-aware, and task-aware attention functions, respectively. The symbol • denotes element-wise feature reweighting.

The scale-aware module (π_L) evaluates the relative importance of feature maps at different pyramid levels. Since tobacco shreds usually appear as slender and small targets with large-scale variations, assigning adaptive weights to different feature levels helps the network preserve both high-resolution spatial details and high-level semantic information. The scale-aware attention is calculated as

$$\pi_L(F)\cdot F = \sigma\left(f\left(\frac{1}{SC}\sum_{s=1}^{S}\sum_{c=1}^{C}F_{s,c}\right)\right)\cdot F\,, \tag{4}$$

where S and C denote the spatial and channel dimensions of the feature tensor, respectively; $F_{s,c}$ represents the feature response at spatial position s and channel c; $f(\cdot)$ is a linear transformation implemented by a 1 × 1 convolution; and $\sigma(\cdot)$ denotes the Hard-Sigmoid activation function used to generate normalized attention weights.

The spatial aware module (π_S) leverages deformable convolution to learn sparse sampling locations in the feature space, thereby better capturing irregular shapes and delicate structures. By aggregating spatial features across different hierarchical levels, the module significantly enhances adaptability to complex geometric forms. Its mathematical formulation is expressed as

$$\pi_S(F)\cdot F = \frac{1}{L}\sum_{l=1}^{L}\sum_{k=1}^{K}w_{l,k}\cdot F\left(l;p_k+\Delta p_k;c\right)\cdot\Delta m_k\,. \tag{5}$$

Here, L represents the total number of feature levels, K denotes the number of sparse sampling points, p_k indicates the initial position coordinates, Δp_k signifies the spatial offset learned by the network to adjust the sampling location, and $w_{l,k}$ corresponds to the weighting coefficient at the respective position.

The task-aware module (π_c) dynamically modulates the activation states of individual channels, enabling the model to adaptively select features based on the requirements of different detection tasks. This approach effectively enhances the task relevance and discriminative power of the feature representations. Its computational procedure is defined as

$$\pi_C(F)\cdot F = \max\left(\alpha^1(F)\cdot F_c+\beta^1(F),\,\alpha^2(F)\cdot F_c+\beta^2(F)\right), \tag{6}$$

where F_c denotes the feature slice of the c-th channel, and the parameter set $\theta = \left[\alpha^1, \alpha^2, \beta^1, \beta^2\right]^T$ consists of a series of hyperfunctions that collectively enable the dynamic regulation of the channel activation thresholds.

This dimension-wise attention architecture not only enhances the model's representational capacity for multi-scale and multi-shape targets but also improves the discriminative power and robustness of the overall features by explicitly modeling interdependencies across dimensions.

2.4 SDB Loss

In object detection models, the precise regression of small targets remains a significant challenge. The inherent ambiguity in their annotated boundaries and low resolution often leads to severe fluctuations in Intersection over Union (IoU)-based loss functions, substantially compromising training stability and final model performance.[15] To address this issue, we introduce SDB Loss in this paper. This loss function incorporates a dynamic modulation factor adaptive to the target scale within the CIoU Loss framework, effectively rebalancing the weight allocation between the center-distance loss and the shape-loss terms, thereby significantly enhancing regression robustness and accuracy for small targets.

SDB Loss is structured around two fundamental loss terms: the Shape Loss term (L_{BS}) and the Center-Distance Loss term (L_{BL}), which are defined as

$$L_{BS} = 1 - IoU + \alpha v, \quad L_{BL} = \frac{\rho^2\left(b_p, b_{gt}\right)}{c^2}, \tag{7}$$

where IoU represents the Intersection over Union between the predicted and ground truth bounding boxes, αv measures the aspect ratio consistency, and ρ denotes the Euclidean distance. The core innovation of the loss function lies in the introduction of a scale-aware dynamic modulation factor β_B, which is calculated as

$$\beta_B = \min\left(\frac{B_{gt}}{B_{gt\max}} \times R_{OC} \times \delta, \delta\right). \tag{8}$$

B_{gt} denotes the area of the ground truth bounding box, $B_{gt\max}$ represents the maximum size of infrared small targets as defined by the International Society for Optics and Photonics (SPIE),[16] and δ is the upper-bound hyperparameter. This factor ensures that β_B is positively correlated with the target size and dynamically allocates weights to the two loss terms accordingly. The final form of the loss function is expressed as

$$\beta_{L_{BS}} = 1 - \delta + \beta_B, \quad \beta_{L_{BL}} = 1 + \delta - \beta_B, \tag{9}$$

$$L_{SDB} = \beta_{L_{BS}} \times L_{BS} + \beta_{L_{BL}} \times L_{BL}\,, \tag{10}$$

where $\beta_{L_{BS}}$ and $\beta_{L_{BL}}$ are the impact factors for L_{BS} and L_{BL}, respectively. When the area of the target bounding box exceeds 81, L_{SDB} degenerates into CIoU Loss.

This mechanism implements a scale-aware optimization strategy: for small targets (with smaller β_{B} values), the model increases the weight of the center-distance loss (increasing $\beta_{L_{BL}}$) and relaxes the shape constraint, thereby alleviating the sensitivity of IoU to minor positional deviations and enhancing training stability; for large targets (where $\beta_{\mathrm{B}} = \delta$), the loss reduces to the standard CIoU form, preserving comprehensive regression performance. By employing this scale-aware dynamic weight allocation mechanism, SDB Loss achieves differentiated optimization for targets of varying sizes, significantly enhancing the model's detection capability and regression stability for small targets in complex scenarios.

3. Dataset Construction and Model Training

3.1 Dataset acquisition

The dataset images were collected from the primary processing line at Xiamen Tobacco Industrial Co., Ltd., where the standard width of the produced tobacco shreds is 0.90 mm. The tobacco shred image acquisition setup consists of an S20 collaborative robot, a vibration chute, a flexible vibration tray, a vision inspection unit, and a glass press plate. The automated tobacco sampling procedure is as follows: after the production line conveyor belt is activated, the host computer triggers the robotic arm according to the programmed sequence. The arm uses an end-effector blower to suction a tobacco shred sample and transfers it to the vibration chute. A graded diversion mechanism then directs shreds of qualified size into the flexible vibration tray. The flexible vibration tray disperses the shreds evenly through vibration. The robot subsequently places the glass press plate onto the shreds to flatten them. Finally, the camera captures and saves the image. A sample image of the collected tobacco shreds is shown in Fig. 6.

3.2 Dataset construction

A total of 1034 original tobacco shred images were acquired using the image acquisition setup described above. The images were annotated using the LabelImg tool in the PyCharm environment, and the annotation files were saved in YOLO-format TXT files. In this study, only tobacco shred segments that were suitable for subsequent width measurement were annotated as target objects. Specifically, a tobacco shred was considered suitable when it was morphologically complete, relatively straight, and individually separable from neighboring shreds, and had clearly visible boundaries and endpoints. Shreds with severe overlap, obvious curling, fragmentation, unclear contours, or incomplete visible structures were not annotated as target objects because they could not provide reliable geometric information for width measurement.

Fig. 6. (Color online) Tobacco shred sample image.

Objects that did not meet the above criteria, including overlapped shreds, debris, excessively short fragments, severely curled shreds, and indistinguishable clustered regions, were treated as background during model training. This annotation strategy allowed the detector to learn not only the appearance of measurable tobacco shreds but also the distinction between measurable and non-measurable samples in complex production-line images.

To avoid data leakage between the training, validation, and test sets, the 1034 original images were first divided into training, validation, and test sets according to an 8:1:1 ratio. Data augmentation was then applied only to the training set. The augmentation operations included rotation, horizontal or vertical mirroring, and random brightness adjustment, which were used to improve model robustness to changes in shred orientation and illumination. After augmentation, the final dataset contained 4134 images. The validation and test sets were kept independent and were not augmented, ensuring that model performance was evaluated on images not derived from the training samples.

The annotation was first performed by one researcher and then reviewed by two enterprise engineers with experience in tobacco production and quality inspection. Ambiguous samples, such as partially overlapped or slightly curled tobacco shreds, were jointly reviewed according to whether their boundaries and centerline can support reliable width measurement. Samples that did not satisfy this requirement were excluded from target annotation and retained as background.

3.3 Model training environment and parameter configuration

The experimental platform was constructed using the Windows 11 operating system and the PyTorch deep learning framework, with Python 3.11, PyTorch 2.2.2, and CUDA 12.3. The hardware configuration comprised an AMD Ryzen 5 9600X CPU and an NVIDIA GeForce RTX 4070 graphics card (12GB). The hyperparameter settings for model training are detailed in Table 1.

Table 1
Model training hyperparameters.

Hyperparameter	Value
Image input size	640 × 640
Batch size	16
Epoch	300
Optimizer	SGD
Learning rate	0.01
Learning rate decay factor	0.01
Momentum	0.937

4. Experimental Results and Analysis

4.1 Model evaluation metrics

On the basis of the actual requirements for tobacco shred width detection, the performance of the trained detection model was evaluated using the following objective metrics: precision (*P*), recall (*R*), mean average precision (*mAP*), *F*1-*score*, and frames per second (*FPS*). The formulas for calculating *P*, *R*, *mAP*, and *F*1-*score* are as follows.

$$P = \frac{TP}{FP + TP} \tag{11}$$

$$R = \frac{TP}{FN + TP} \tag{12}$$

$$AP = \int_0^1 P(R)\,dR \tag{13}$$

$$mAP = \frac{\sum_{i=1}^{n} AP_i}{n} \tag{14}$$

$$F1\text{-}score = \frac{2 \times R \times P}{R + P} \tag{15}$$

In the formulas, *TP* represents the number of instances where the ground truth is positive and the prediction is also positive, *TN* represents the number of instances where the ground truth is negative and the prediction is also negative, *FP* represents the number of instances where the ground truth is negative but the prediction is positive, and *FN* represents the number of instances where the ground truth is positive but the prediction is negative.

P and *R* typically exhibit a trade-off relationship;[17] when *R* is high, *P* tends to be relatively low, and conversely, high *P* often corresponds to low *R*. To balance *P* and *R* and provide a more comprehensive evaluation of model performance, average precision (*AP*) is more suitable. *AP* is

defined as the area under the *P*–*R* curve, plotted with *R* on the horizontal axis and *P* on the vertical axis. In this single-class detection task, *mAP* is equivalent to the *AP* of the tobacco-shred class under the specified IoU threshold. *F*1-*score* represents the harmonic mean of *P* and *R*, with values ranging from 0 to 1. A higher *F*1-*score* indicates higher model performance, provides a more balanced assessment, and is often used as a final evaluation metric.[18]

4.2 Ablation study analysis

To validate the effectiveness of the proposed improvements, ablation studies were conducted using YOLO11n as the baseline model by systematically evaluating the three modified modules. The detailed training results for each model configuration are presented in Table 2, and the *P*–*R* curves for the different models are shown in Fig. 7.

The ablation study results in Table 2 show the effects of the three improved components on detection accuracy. Compared with the baseline YOLO11n model, the independent introduction

Table 2
Ablation study results.

Model	BiFPN	Dynamic Head	SDB Loss	*P* (%)	*R* (%)	*mAP* (%)	*F*1-*score*
1				70.75	72.09	75.46	0.7141
2	✓			69.66	71.35	74.70	0.7049
3		✓		80.26	73.73	83.59	0.7686
4			✓	73.65	70.47	77.00	0.7203
5	✓	✓		78.26	77.52	84.55	0.7789
6	✓		✓	71.29	72.21	75.63	0.7175
7		✓	✓	80.17	76.86	84.85	0.7848
8	✓	✓	✓	83.02	76.00	86.61	0.7935

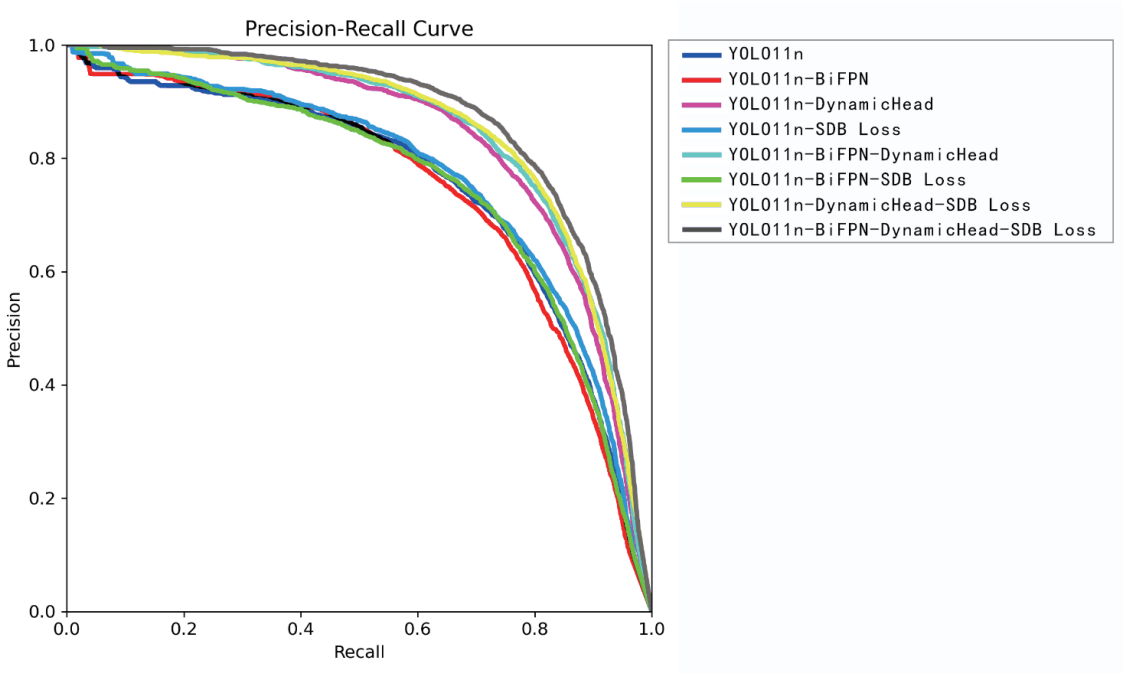

Fig. 7. (Color online) *P*–*R* curves for different models.

of BiFPN slightly decreased *mAP* from 75.46 to 74.70% and reduced *F*1-*score* from 0.7141 to 0.7049. This indicates that BiFPN alone does not bring a clear accuracy gain in this task. The individual incorporation of Dynamic Head substantially improved *mAP* to 83.59% and increased *F*1-*score* to 0.7686, showing its effectiveness in enhancing feature representation for slender and irregular tobacco shreds. The standalone use of SDB Loss improved *mAP* to 77.00% and *F*1-*score* to 0.7203, suggesting that the scale-aware regression loss helps improve localization performance for small targets. In the combination experiments, the joint use of BiFPN and Dynamic Head achieved an *mAP* of 84.55% and an *F*1-*score* of 0.7789. The combination of Dynamic Head and SDB Loss further increased *mAP* to 84.85% and *F*1-*score* to 0.7848. The complete model integrating BiFPN, Dynamic Head, and SDB Loss achieved the highest detection accuracy, with an *mAP* of 86.61% and an *F*1-*score* of 0.7935. Compared with the original YOLO11n model, the complete model improved *mAP* by 11.15 percentage points and increased *F*1-*score* by 0.0794, demonstrating the effectiveness of the proposed combined improvements.

To further illustrate the practical effects of the proposed improvements, Fig. 8 shows the detection results of the eight ablation models on the same tobacco shred image. The baseline YOLO11n model detected 33 targets [Fig. 8(a)], but several boxes corresponded to repeated responses on the same shred or to highly adjacent fragments. After introducing BiFPN, the model detected 34 targets [Fig. 8(b)], indicating that multi-scale feature fusion helped preserve small-shred responses, but some detections still appeared on clustered or less suitable fragments. In contrast, the Dynamic Head module reduced the number of detections to 27 [Fig. 8(c)], suggesting that task-aware feature representation helped suppress part of the redundant detections. The model trained with SDB Loss detected 31 targets [Fig. 8(d)], with the detected boxes showing improved consistency with measurable shred regions.

The comparison of combined modules further demonstrates their complementary effects. The YOLO11n + Dynamic Head + BiFPN model detected 31 targets [Fig. 8(e)], preserving multi-scale detection ability while reducing some repeated responses. The YOLO11n + BiFPN + SDB Loss model detected 35 targets [Fig. 8(f)], showing strong target retention but still including several adjacent or potentially unsuitable fragments. When Dynamic Head and SDB Loss were combined, the number of detections decreased to 26 [Fig. 8(g)], indicating a stronger suppression of unstable or redundant detections. Finally, the complete model integrating Dynamic Head, BiFPN, and SDB Loss produced a more balanced detection result [Fig. 8(h)], retaining representative isolated tobacco shreds while reducing the numbers of repeated detections and detections on overlapped or curled fragments. Therefore, for the present task, a larger number of detection boxes does not necessarily indicate higher performance, because the detector is intended to select isolated and morphologically suitable tobacco shreds for subsequent width measurement rather than identify all visible fragments. This visual evidence is consistent with the quantitative results in Table 2, where the complete model achieves the highest *mAP* of 86.61% and the highest *F*1-*score* of 0.7935.

To further clarify the relationship between model complexity and inference speed, Table 3 shows the number of parameters, FLOPs, model size, total latency, and *FPS* for each model. The BiFPN-only model slightly increased the number of parameters from 2.590 to 2.614 M and

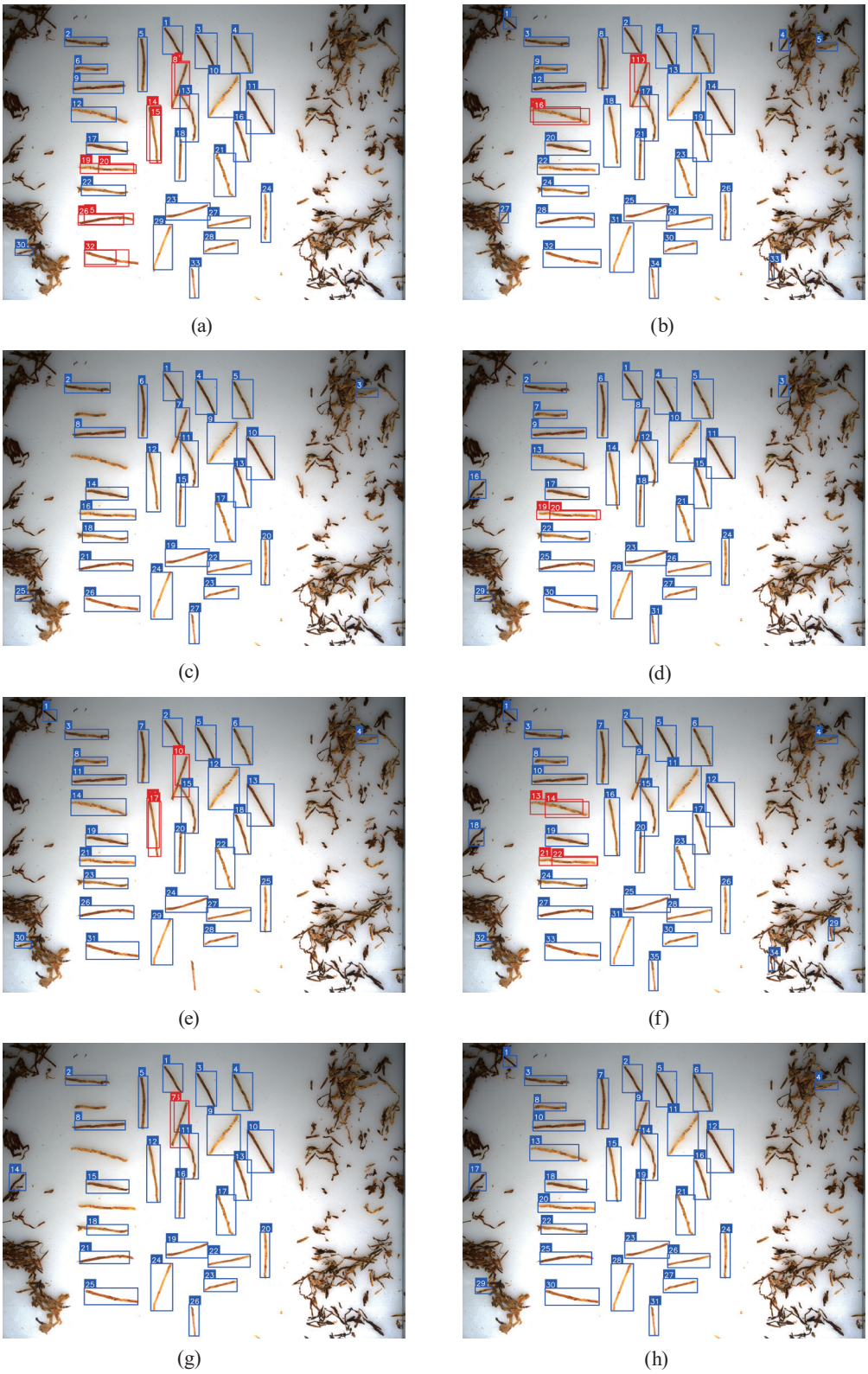

Fig. 8. (Color online) Visual comparison of detection results from the eight ablation models on the same tobacco shred image. (a) YOLO11n. (b) YOLO11n + BiFPN. (c) YOLO11n + Dynamic Head. (d) YOLO11n + SDB Loss. (e) YOLO11n + Dynamic Head + BiFPN. (f) YOLO11n + BiFPN + SDB Loss. (g) YOLO11n + Dynamic Head + SDB Loss. (h) YOLO11n + Dynamic Head + BiFPN + SDB Loss.

Table 3
Model complexity and runtime comparison.

Model	No. of parameters (M)	FLOPs (G)	Model size (MB)	Total latency (ms·image^{-1})	*FPS* (f·s^{-1})
1	2.590	6.441	5.238	20.09 ± 2.18	50.34 ± 5.23
2	2.614	7.072	5.342	24.84 ± 2.40	40.61 ± 3.82
3	3.941	9.583	7.847	46.98 ± 3.47	21.40 ± 1.55
4	2.590	6.441	5.238	19.98 ± 2.12	50.58 ± 5.21
5	3.867	10.397	7.763	50.78 ± 3.78	19.80 ± 1.44
6	2.614	7.072	5.342	24.98 ± 2.06	40.31 ± 3.31
7	3.941	9.583	7.847	46.89 ± 3.85	21.47 ± 1.77
8	3.867	10.397	7.763	52.65 ± 5.40	19.19 ± 1.92

FLOPs from 6.441 to 7.072 G, resulting in a lower *FPS* than the baseline. The Dynamic-Head-based models introduced more parameters and FLOPs, which explains their higher latency and lower *FPS*. The SDB-Loss-only model had the same parameters, FLOPs, and model size as the baseline because SDB Loss is used only during training and does not change the inference architecture. Therefore, its contribution should be interpreted mainly as improving localization accuracy rather than directly accelerating inference.

The complete model achieved 19.19 ± 1.92 *FPS* with a total latency of 52.65 ± 5.40 ms per image. Although its inference speed was lower than that of the baseline model, it provided the highest detection accuracy and remained suitable for tobacco shred sampling detection. As shown by the *P*–*R* curves in Fig. 7, the complete model obtained the largest curve area among the compared variants, further confirming the effectiveness of the proposed improvements.

4.3 Comparative experiment analysis

To further validate the performance advantages of the proposed improved model in tobacco shred detection, comparative experiments were conducted with Faster R-CNN, YOLOv5n, YOLOv8n, YOLOv10n, YOLO11n, and our enhanced model. All experimental data are presented in Table 4.

The comparative results show that the proposed model achieved the highest overall performance among the tested detection models. Compared with Faster R-CNN, YOLOv5n, YOLOv8n, YOLOv10n, and the original YOLO11n, the proposed model obtained higher *P*, *mAP*, and *F*1-*score*. Specifically, *P* increased by 10.67–12.76 percentage points, *mAP* increased by 11.15–14.30 percentage points, and *F*1-*score* increased by 0.0748–0.0873. These results indicate that the proposed improvements enhance the model's ability to select measurable tobacco shreds while reducing the number of missed and false detections under complex production-line imaging conditions.

5. Tobacco Shred Width Measurement

5.1 Tobacco shred width calculation

The objective of this study on tobacco shred detection algorithms is to identify and localize suitable shred segments within images, although it does not enable the quantitative analysis of

Table 4
Comparative experiment results.

Model	*P* (%)	*R* (%)	*mAP* (%)	*F*1-*score*
Faster R-CNN	71.34	69.92	74.06	0.7062
YOLOv5n	71.94	70.52	72.31	0.7122
YOLOv8n	70.26	71.66	75.32	0.7095
YOLOv10n	72.35	71.40	74.38	0.7187
YOLO11n	70.75	72.09	75.46	0.7141
Model in this paper	83.02	76.00	86.61	0.7935

width values. Therefore, after extracting the targets, it is necessary to incorporate a width measurement algorithm to determine specific width values. The procedure first uses the improved YOLO11n algorithm to localize and identify tobacco shreds, determine the precise positions of target shred segments, and extract their corresponding rectangular regions. Subsequently, the variable-diameter circle method is applied to measure the shred width, as schematically illustrated in Fig. 9.

To improve the detection accuracy of tobacco shred width, the separated shred images undergo preprocessing steps, including grayscale conversion, binarization, and morphological operations such as opening, closing, and noise reduction, to enhance image features and facilitate subsequent analysis.

In this study, we employed the maximum diameter of variable-diameter circles to characterize the local width of tobacco shreds. First, the Zhang-Suen thinning algorithm[19] is applied to extract the shred centerline, and its length N is measured. Subsequently, starting from one end of the centerline, a series of points is selected at intervals of N/10 to serve as the centers of the variable-diameter circles. With an initial radius of 1 pixel, the circle's radius is incrementally increased while its center position and radius are dynamically adjusted on the basis of the circle's intersection with the contour. This process continues until the circle becomes tangent to the contour, at which point the current circle's diameter is recorded as the local shred width at that position. After iterating through all center points along the centerline, the average of the local width measurements is calculated as the final width measurement for the tobacco shred. The complete width-measurement workflow, including target-region extraction, image preprocessing, centerline extraction, and variable-radius-circle fitting, is shown in Fig. 10.

5.2 Validation of measurement accuracy

To validate the complete visual selection and width measurement pipeline under production-batch conditions, field validation data from three tobacco shred batches were used. The tested batches included Qipilang Chunjing, Qipilang Blue, and Qipilang Gutian. For each batch, 10 sampling tests were conducted, and each sampling test included 10 tobacco shreds. Therefore, the validation covered 30 sampling tests and 300 tobacco shreds in total. The manual reference values were obtained by professional enterprise quality inspectors using the standard tobacco projection measurement method. The relative error was calculated as the absolute difference between the manual reference value and the software measurement value divided by the manual reference value.

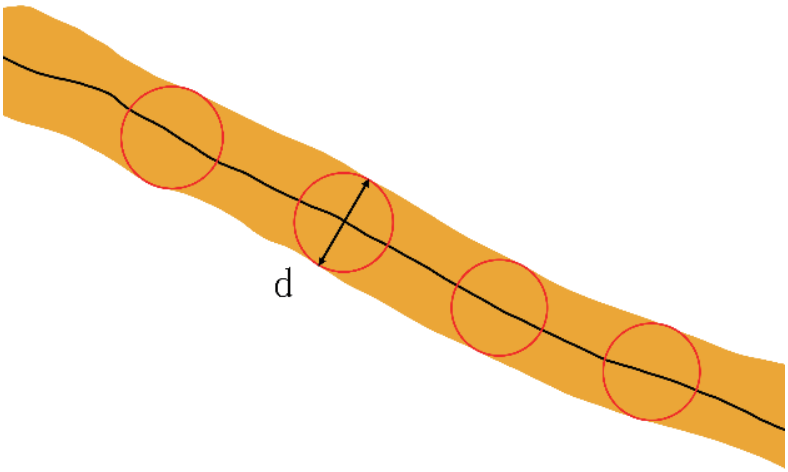

Fig. 9. (Color online) Schematic diagram of width measurement using variable-diameter circle method.

Table 5
Batch-level validation results of tobacco shred width measurement.

Batch	Sampling tests	No. of shreds per test	Total no. of shreds	Mean relative error/%	Maximum relative error/%	Enterprise acceptance threshold/%
Qipilang Chunjing	10	10	100	4.37	7.25	8.00
Qipilang Blue	10	10	100	5.52	6.94	8.00
Qipilang Gutian	10	10	100	4.10	6.95	8.00
Overall	30	10	300	4.66	7.25	8.00

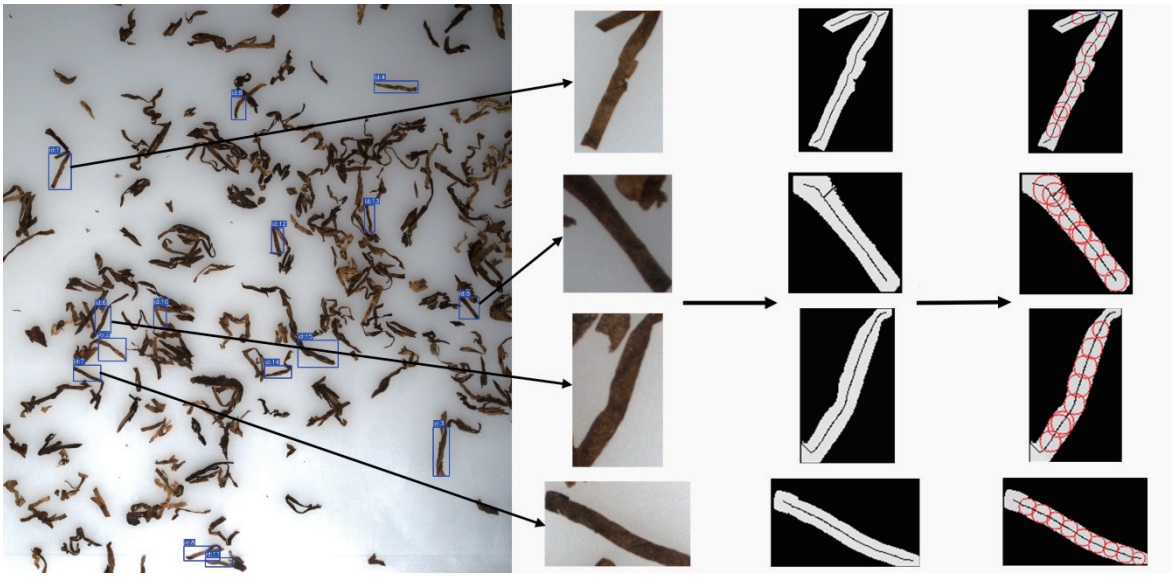

Fig. 10. (Color online) Tobacco shred width measurement process.

As shown in Table 5, the average relative errors of Qipilang Chunjing, Qipilang Blue, and Qipilang Gutian were 4.37, 5.52, and 4.10%, respectively. The maximum relative errors of the three batches were 7.25, 6.94, and 6.95%, respectively. Across all 30 sampling tests, the overall average relative error was 4.66% and the maximum relative error was 7.25%. Since all sampling-level errors were lower than the enterprise acceptance threshold of 8%, the proposed measurement pipeline fully satisfied the enterprise requirements for tobacco shred width inspection under the tested production-batch conditions. The expanded batch-level validation provides stronger evidence for the practical applicability of the proposed method.

6. Conclusion

In this study, we proposed a YOLO11n-based visual selection and width measurement pipeline for shredded tobacco under complex production-line imaging conditions. By integrating BiFPN, Dynamic Head, and SDB Loss, the improved model achieved an *mAP* of 86.61% and an *F*1-*score* of 0.7935, improving *mAP* by 11.15 percentage points and *F*1-*score* by 0.0794 compared with the original YOLO11n model. Under the unified runtime test, the complete model achieved 19.19 ± 1.92 *FPS*, which is acceptable for sampling-based tobacco shred width inspection.

The selected tobacco shred regions were further processed using a variable-radius circle method for width estimation. Field validation on three production batches, covering 30 sampling tests and 300 tobacco shreds, showed an overall average relative error of 4.66% and a maximum relative error of 7.25%. All sampling-level errors were below the enterprise acceptance threshold of 8%, indicating that the proposed pipeline fully satisfied the enterprise requirements under the tested production-batch conditions. Future work will focus on broader production-condition validation and lightweight deployment on industrial computers.

Acknowledgments

This work was supported by the Science and Technology Project of Xiamen Tobacco Industry Co., Ltd. (FJZYKJJH2023ZD055) and the Fujian Provincial Science and Technology Major Project of China (2024HZ026020).

References

1 Y. Zhao, T. Zhang, X. Qiao, J. Zhao, Q. Hu, H. Feng, L. Yang, J. Wu, X. He, Q. Pu, M. Zhao, and Y. Qiu: J. Phys.: Conf. Ser. **1748** (2021) 062060. https://doi.org/10.1088/1742-6596/1748/6/062060
2 L. Wang, S. Xu, Z. Tao, and Q. Zhang: Proc. 3rd Int. Conf. Cognitive Based Information Processing and Applications (Springer, Singapore, 2024) 117–127. https://doi.org/10.1007/978-981-97-1983-9_11
3 W. Zhao, R. Wen, J. Zeng, C. Di, W. Yan, D. Li, M. Liu, and Y. Liu: Tob. Sci. Technol. **46** (2013) 12. https://doi.org/10.3969/j.issn.1002-0861.2013.10.003
4 H. Liu, G. Deng, B. Li, B. Zhou, J. Zhao, Z. Chen, and W. Zhu: Tob. Sci. Technol. **55** (2022) 77. https://doi.org/10.16135/j.issn1002-0861.2021.0047
5 S. Hu, J. Chen, D. Lai, and L. Zhou: J. Comput. Appl. **39** (2019) 215.
6 X. Liu, Z. Xie, W. Guan, R. Chen, and S. Kuang: Mach. Des. Manuf. Eng. **53** (2024) 127. https://doi.org/10.3969/j.issn.2095-509X.2024.06.025
7 J. Zhao and Y. Hu: Comput. Appl. Softw. **42** (2025) 63. https://doi.org/10.3969/j.issn.1000-386x.2025.04.011

8 L. He, H. Wei, and Q. Wang: Sensors **23** (2023) 6477. https://doi.org/10.3390/s23146477
9 X. Dai, Y. Chen, B. Xiao, D. Chen, M. Liu, L. Yuan, and L. Zhang: Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (IEEE, 2021) 7369–7378. https://doi.org/10.1109/CVPR46437.2021.00729
10 J. Yang, S. Liu, J. Wu, X. Su, N. Hai, and X. Huang: Proc. AAAI Conf. Artif. Intell. **39** (2025) 9202. https://doi.org/10.1609/aaai.v39i9.32996
11 Z. Luo: Proc. 2025 8th Int. Conf. Advanced Algorithms and Control Engineering (IEEE, 2025) 1324–1327. https://doi.org/10.1109/ICAACE65325.2025.11020323
12 S. Wang, Q. Dong, X. Chen, Z. Chu, R. Li, J. Hu, and X. Gu: J. Infrastruct. Syst. **30** (2024) 04024005. https://doi.org/10.1061/JITSE4.ISENG-2389
13 N. Li, T. Ye, Z. Zhou, C. Gao, and P. Zhang: Appl. Sci. **14** (2024) 429. https://doi.org/10.3390/app14010429
14 L. Deng, Z. Miao, X. Zhao, S. Yang, Y. Gao, C. Zhai, and C. Zhao: Agronomy **15** (2025) 57. https://doi.org/10.3390/agronomy15010057
15 A. Ahmad, Ed.: Handbook of Optomechanical Engineering (CRC Press, Boca Raton, 2017) 2nd ed. https://doi.org/10.4324/9781315153247
16 T. Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár: Proc. IEEE Int. Conf. Computer Vision (IEEE, 2017) 2999–3007. https://doi.org/10.1109/ICCV.2017.324
17 Y. He, C. Zhu, J. Wang, M. Savvides, and X. Zhang: Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition (IEEE, 2019) 2883–2892. https://doi.org/10.1109/CVPR.2019.00300
18 S. J. Jiao and Z. Y. Guo: Meas. Control Technol. **44** (2025) 35. https://doi.org/10.19708/j.ckjs.2025.04.302
19 T. Y. Zhang and C. Y. Suen: Commun. ACM **27** (1984) 236. https://doi.org/10.1145/357994.358023