



*Research article*

## **CJS-YOLOv5n: A high-performance detection model for cigarette appearance defects**

**Yihai Ma<sup>1,2</sup>, Guowu Yuan<sup>1,2,\*</sup>, Kun Yue<sup>1,2</sup> and Hao Zhou<sup>1,2</sup>**

<sup>1</sup> School of Information Science and Engineering, Yunnan University, Kunming 650504, China

<sup>2</sup> Yunnan Key Laboratory of Intelligent Systems and Computing, Kunming 650504, China

\* **Correspondence:** E-mail: [gwyuan@ynu.edu.cn](mailto:gwyuan@ynu.edu.cn); Tel: +8687165033748.

**Abstract:** In tobacco production, cigarettes with appearance defects are inevitable and dramatically impact the quality of tobacco products. Currently, available methods do not balance the tension between detection accuracy and speed. To achieve accurate detection on a cigarette production line with the rate of 200 cigarettes per second, we propose a defect detection model for cigarette appearance based on YOLOv5n (You Only Look Once Version 5 Nano), called CJS-YOLOv5n (YOLOv5n with C2F (Cross Stage Partial (CSP) Bottleneck with 2 convolutions-fast), Jump Concat, and SCYLLA-IoU (SIoU)). This model incorporates the C2F module proposed in the state-of-the-art object detection network YOLOv8 (You Only Look Once Version 8). This module optimizes the network by parallelizing additional gradient flow branches, enhancing the model's feature extraction capability and obtaining richer gradient information. Furthermore, this model uses Jump Concat to preserve minor defect feature information during the fusion process in the feature fusion pyramid's P4 layer. Additionally, this model integrates the SIoU localization loss function to improve localization accuracy and detection precision. Experimental results demonstrate that our proposed CJS-YOLOv5n model achieves superior overall performance. It maintains a detection speed of over 500 FPS (frames per second) while increasing the recall rate by 2.3% and mAP (mean average precision)<sub>@0.5</sub> by 1.7%. The proposed model is suitable for application in high-speed cigarette production lines.

**Keywords:** cigarette; defect detection; YOLOv5n; C2F; Jump Concat; SIoU

---

## 1. Introduction

The quality of tobacco products is an essential indicator of tobacco enterprises' production level and brand establishment. Cigarettes inevitably have appearance defects on the production lines, such as stains, scratches and wrinkles. There is an urgent need for tobacco companies to adopt strict quality control measures to prevent cosmetically defective products from entering the sales market. In the past, manual inspection was the main method for cigarette factories to detect appearance defects. However, this method has problems such as strong subjectivity, slow detection speed, low efficiency and high missed detection rate, and it cannot adapt to the high-speed and high-precision detection requirements of modern industry. Therefore, there is an urgent need to develop an efficient, accurate and cost-effective automated detection method to replace manual detection.

Product appearance defect detection based on conventional computer vision algorithms often relies on template matching to identify defect locations, entailing complications in template design and yielding suboptimal accuracy. With the advancements in deep learning technology, deep learning methods have also been harnessed for product appearance defect detection, resulting in enhanced detection accuracy.

Product appearance defect detection represents a specific application of object detection models. Based on deep learning, object detection models can be broadly categorized into two types: two-stage detectors and one-stage detectors. Two-stage detectors, exemplified by region-based convolutional neural networks (RCNN) [1] and Faster-RCNN [2,3], and one-stage detectors, represented by You Only Look Once (YOLO) [4–7] and Single Shot MultiBox Detector (SSD) [8], have been employed in product appearance defect detection. For instance, Liu et al. [9] utilized an improved YOLOv5 model to detect surface defects on diode metal bases, incorporating attention mechanisms and the K-means++ clustering algorithm, achieving  $mAP@0.5$  84.0%. Chen et al. [10] applied an improved YOLOv3 model to detect surface defects on surface-mounted device (SMD) LED chips, replacing the backbone network with DenseNet and achieving  $mAP@0.5$  95.28%. Hu and Wang [11] employed an improved Faster R-CNN model to detect surface defects on printed circuit boards (PCB), replacing the model's backbone network and feature fusion pyramid, achieving  $mAP@0.5$  94.2%. Li and Xu [12] used an improved SSD model for appearance defect detection on electronic products, replacing the backbone network with MobileNet and optimizing parameters to achieve  $mAP@0.5$  88.6%. Duan et al. [13] utilized an improved YOLOv3 model for surface defect detection in casting, introducing dual-density convolutions and low-resolution detection layers to achieve  $mAP@0.5$  88.02%. Kou et al. [14] employed an improved YOLOv3 model for surface defect detection on steel strips, modifying the detection head to Anchor-Free and achieving  $mAP@0.5$  71.3% on the GC10-DET dataset.

In defect detection research for cigarette appearance, Yuan et al. [15] employed an improved ResNeSt model, achieving a classification accuracy of 94.01% through transfer learning, multi-scale learning methods and modified activation functions. Qu et al. [16] utilized an improved SSD model for cigarette appearance defect detection, achieving  $mAP@0.5$  90.26% by improving the pyramid convolution. However, the detection speed still fell short of 100 frames per second (FPS). Liu and Yuan [17] employed an improved YOLOv5s model for cigarette appearance defect detection, achieving  $mAP@0.5$  94.0% by introducing attention mechanisms and modifications to the loss function. Nonetheless, the recall rate reached only 86.8%. Yuan et al. [18] employed an improved YOLOv4 model for cigarette appearance defect detection, achieving  $mAP@0.5$  91.46% through introducing attention mechanisms, dilated pyramid spatial pooling and modifications to anchor

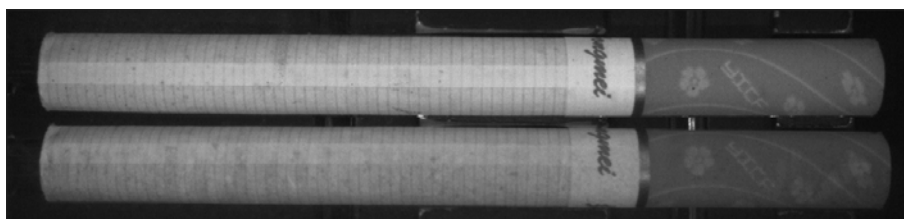
clustering algorithms and loss functions. However, the recall rate remained at 88.81%. Liu et al. [19] utilized an improved CenterNet model for cigarette appearance defect detection, achieving mAP@0.5 95.01% by introducing attention mechanisms, deformable convolutions, feature fusion pyramids and modifications to activation functions and data augmentation methods. Nonetheless, the recall rate reached only 85.96%.

Although there has been considerable research on industrial appearance defect detection using deep learning, studies specifically addressing cigarette appearance defect detection remain limited. Existing research on cigarette appearance defect detection still suffers from low detection recall rates and slow detection speeds. With the advancement of automation in tobacco production, cigarette production lines can now achieve speeds of up to 200 cigarettes per second, posing challenges for real-time detection using existing methods. To address this practical demand, we propose a real-time defect detection method for cigarette appearance based on YOLOv5n. Building upon the YOLOv5 model, our approach incorporates the C2F convolution module, Jump Concat fusion pyramid and SIOU loss function. Experimental results demonstrate that our improved model enhances the precision and recall rate of cigarette appearance defect detection, effectively reducing the presence of defective cigarettes in the consumer market. This model provides robust support for tobacco companies in improving the quality of their cigarette products.

## 2. Dataset

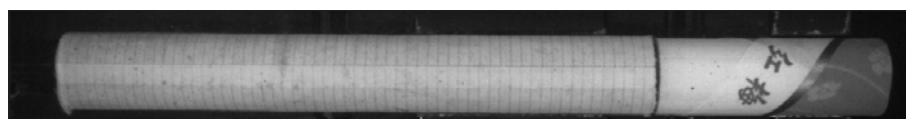
The cigarette images used in our experiments are from the Yunnan Branch of China Tobacco Industry Company Limited. The high-speed industrial cameras capture the images on the automated production line. The front and back cigarettes can be captured at different production line positions. A standard cigarette is 84 mm in length and 7.8 mm in diameter.

Cigarettes are primarily composed of two parts: the cigarette rod, depicted as the white portion on the left side in Figure 1, and the filter, represented by the dark portion on the right side. During cigarette manufacturing, excessive tobacco stems can puncture cigarette rod packaging during rolling [20]. During the twisting process, a certain amount of pressure is applied to ensure a secure bond between the tipping paper and the cigarette rod. However, variations in filter elasticity, roundness of non-filter cigarettes, adhesive properties of latex and the absorptive characteristics of the tipping paper can result in insufficient bonding, misalignment, filter detachment, as well as folding, creasing and misalignment of the tipping paper [21]. Furthermore, improper printing on the cigarette packaging paper or staining during production and transportation can result in visual contamination of the cigarette's appearance.



**Figure 1.** Normal cigarette examples.

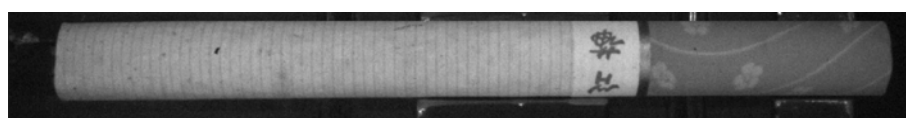
To improve the detection of cigarette appearance defects, we categorized defective cigarettes into seven types: misplacement, stain, scratched, wrinkle, unpacked, no filter and bend, as shown in Figure 2.



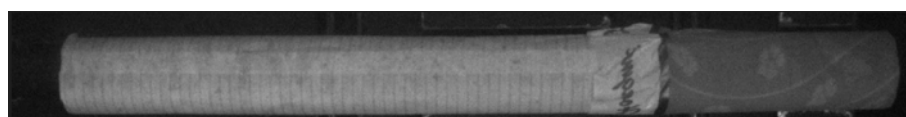
(a) misplacement



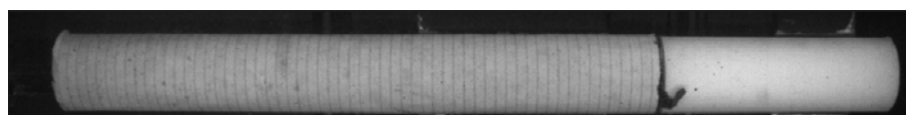
(b) stain



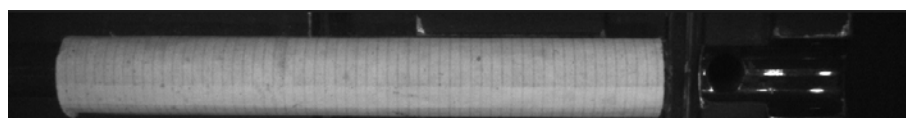
(c) scratched



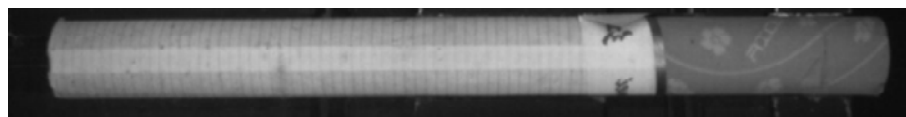
(d) wrinkle



(e) unpacked



(f) no filter



(g) bend

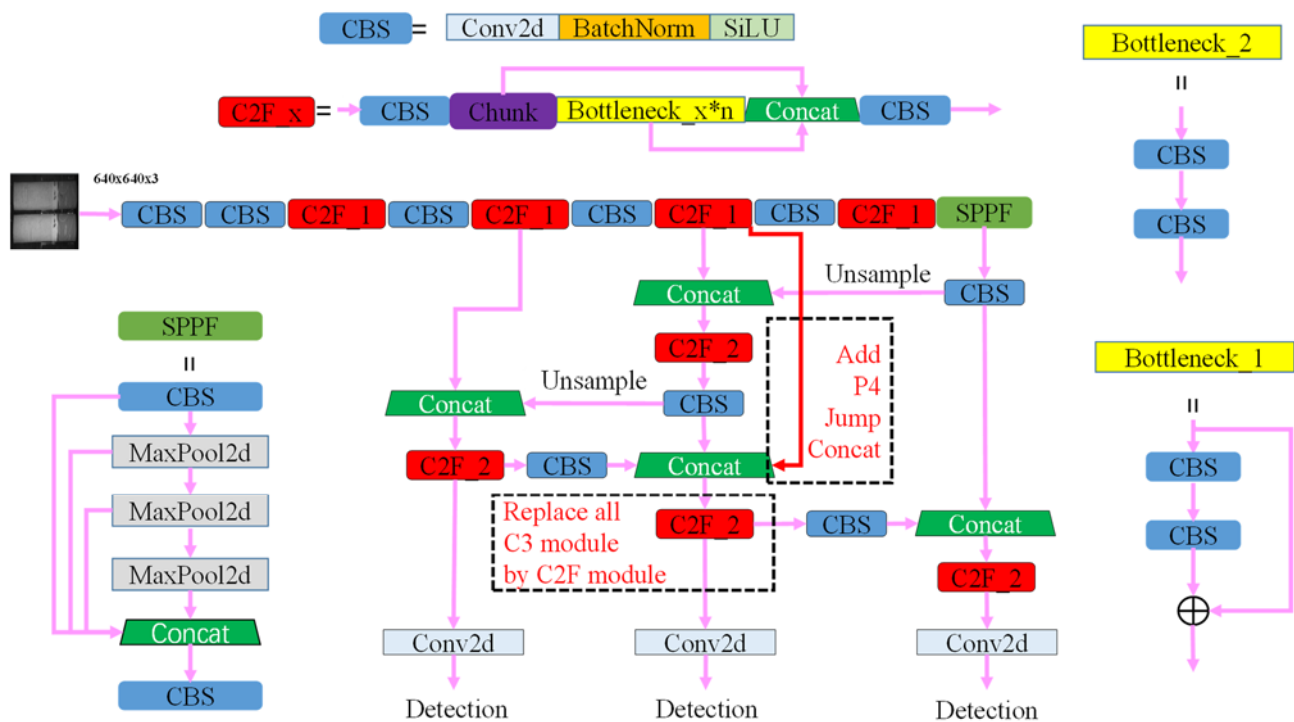
**Figure 2.** Cigarette examples with appearance defects.

### 3. Methods

YOLOv5 comprises five network variants: YOLOv5n, YOLOv5s, YOLOv5m, YOLOv5l and YOLOv5x. These models share the same network architecture but differ in module depth and width.

YOLOv5n and YOLOv5s have the same depth, with YOLOv5n being the shallowest among the YOLOv5 series. To improve the network's runtime on small-scale devices, YOLOv5n reduces the network's width by half compared to YOLOv5s, significantly reducing the scale of parallel computations and enhancing runtime performance on low-capacity devices [22]. To meet the real-time detection requirements and improve recognition speed, we select YOLOv5n as the baseline model.

YOLOv5 comprises four components: input, backbone network, neck network and detection head. The input component incorporates mosaic data augmentation, adaptive anchor calculation and adaptive image scaling. The backbone network performs feature extraction. The neck network employs a top-down and bottom-up feature fusion pathway to merge image features. The detection head includes three layers corresponding to feature maps of sizes  $80 \times 80$ ,  $40 \times 40$  and  $20 \times 20$  for detecting objects of different scales. Finally, the Complete IoU (CIoU [23]) loss function is utilized to calculate the distance between predicted and ground truth bounding boxes and non-maximum suppression (NMS) is applied to remove redundant boxes while retaining those with the highest confidence.



**Figure 3.** CJS-YOLOv5n network structure.

Due to the smaller depth and width of YOLOv5n, its ability to fit complex features is relatively limited. Therefore, this research introduces the C2F module proposed in the state-of-the-art object detection model YOLOv8 [24]. Compared to the original C3 (CSP Bottleneck with 3 convolutions) module [25], the C2F module incorporates bottleneck skip connections, allowing for better feature extraction by increasing the gradient flow and channel capacity. Additionally, only half of the feature matrix participates in subsequent multiple bottleneck operations by splitting the channel number equally after one convolution. This design enhances feature extraction capabilities without compromising GPU inference speed, thus avoiding false detections and omissions.

The original YOLOv5's feature fusion pyramid involves multiple fusion operations on the P4

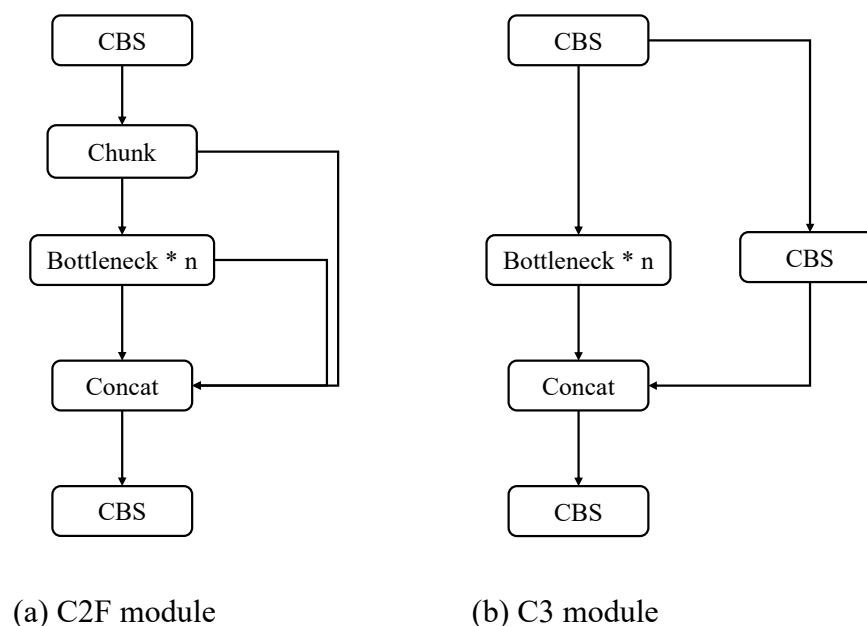
feature layer, which can lead to occlusion and coverage of subtle texture folds in cigarettes, increasing the difficulty of detection. Therefore, this research uses Jump Concat to the P4 feature layer to mitigate the occlusion of fine-textured crease defects resulting from feature map fusion.

The original YOLOv5 utilizes the CIoU localization loss function, which calculates the localization loss based on IoU (Intersection over Union), center point distance and aspect ratio between predicted and ground truth bounding boxes. However, it does not account for mismatched orientations between the predicted and ground truth boxes. This limitation slows down convergence and efficiency, as predicted boxes may “hover” during training, ultimately leading to suboptimal model performance. Therefore, SIoU [26] is introduced, which considers the angle between the center point vectors and includes angle penalty metrics. This loss function enables predicted boxes to converge rapidly towards the nearest axis and converge only on a single axis. This loss function effectively accelerates the convergence of predicted boxes, improving the model’s localization accuracy and confidence in object detection.

Our improved model structure is shown in Figure 3.

### 3.1. C2F module

The C2F module, proposed in the latest object detection model YOLOv8, enhances feature extraction capabilities compared to the original C3 module by introducing bottleneck skip connections and chunk operations, resulting in increased parameter quantity, gradient flow and channel capacity. In Figure 4(a), the C2F module is depicted, while Figure 4(b) represents the original C3 module. The C2F module possesses more channels and gradient flow than the C3 module. The chunk operation performs slicing on the input feature map, dividing it equally into two parts along the channel dimension. Only half of the feature map participates in the bottleneck convolution operation, effectively reducing computational load. Additionally, to maintain a lightweight structure, the C2F module reduces the computation of the right branch CBS (Convolution, Batch normalization and SiLU activation function) module.

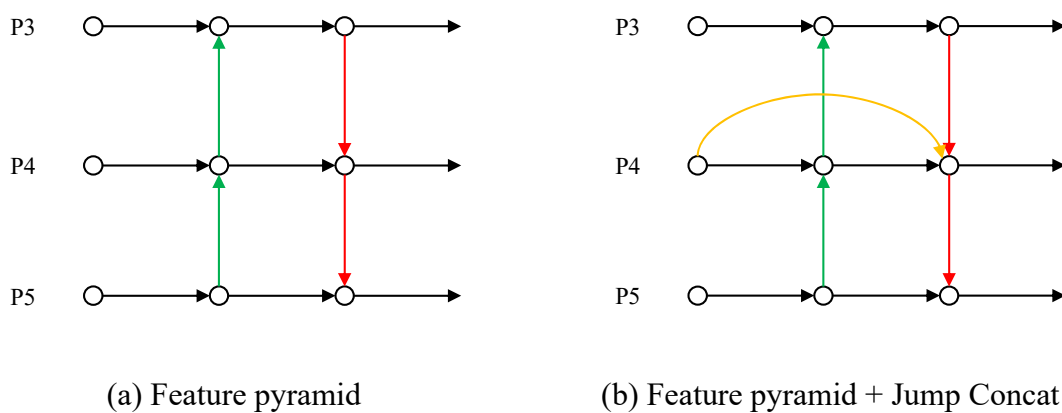


**Figure 4.** Comparison of convolutional module structures.

### 3.2. Jump Concat feature pyramid

The YOLOv5n model incorporates both FPN (Feature Pyramid Network) [27] and PANet (Path Aggregation Network) [28] for multi-scale feature fusion. FPN enhances semantic information in a top-down manner, while PANet reinforces positional information in a bottom-up way. This combination enhances the feature fusion capability of the neck layer. However, in the P4 feature layer, multiple feature fusion operations may cause certain feature information to be overshadowed. To address this issue, this study introduces an additional Jump Concat at the end of the P4 feature layer to prevent the coverage of subtle textures with minimal pixel variations, such as small wrinkles in cigarette appearances.

Figure 5(a) illustrates the original bi-directional feature fusion pyramid of YOLOv5, while Figure 5(b) depicts the enhanced bi-directional feature fusion pyramid with the additional skip connection.



**Figure 5.** Comparison of feature pyramid.

For P4, the two fusion feature processes are formed as follows:

$$P4_{out} = Conv(Resize(P3_{out}) + P4_{in} + Conv(P4_{in} + Resize(Conv(P5_{in})))) \quad (1)$$

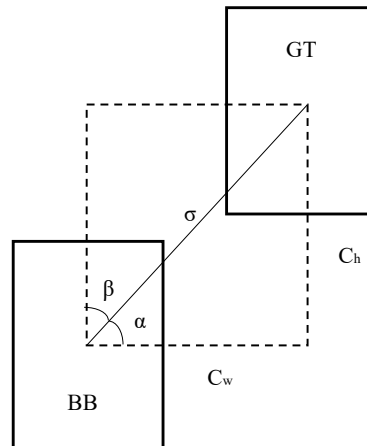
where,  $Resize(\bullet)$  is usually an upsampling or downsampling operation for resolution matching, and  $Conv(\bullet)$  is usually a convolutional operation for feature processing.  $Pi_{out}$  is the output of the  $i$ -th feature map and  $Pi_{in}$  is the input of the  $i$ -th feature map.

### 3.3. SIOU loss

The SIOU loss function is employed for bounding box regression. Compared to CIOU, DIOU (Distance IoU) [29] and GIoU (General IoU) [30], SIOU further considers the vector angle between the ground truth and predicted anchor boxes, leading to a redefinition of the loss function. It facilitates faster regression of anchor boxes towards the nearest axis (x or y), thereby enhancing convergence speed and localization accuracy.

In Figure 6, BB represents the predicted anchor box, GT denotes the ground truth box,  $\sigma$  represents the Euclidean distance between the centers of the predicted and ground truth boxes,  $C_h$  represents the projection distance of the center point on the y-axis,  $C_w$  represents the projection

distance of the center point on the x-axis,  $\alpha$  represents the angle between the center line and the x-axis and  $\beta$  represents the angle between the center line and the y-axis.



**Figure 6.** Principle of SIOU loss function.

The SIOU loss function involves three components in its computation: angle loss, distance loss and intersection over union loss:

$$IoU = \frac{|BB \cap GT|}{|BB \cup GT|} \quad (2)$$

The above equation represents the calculation of the intersection over union, where  $BB$  denotes the area of the predicted bounding box and  $GT$  represents the area of the ground truth box. The intersection ratio to the predicted and ground truth box union yields the  $IoU$  value.

$$\Delta = \sum_{t=x,y} 1 - e^{-\gamma \rho^t} \quad (3)$$

The formula above calculates the angle,  $\Delta$ , which comprises the angle and distance calculations involved in SIOU. It represents the final result of the angle calculation.  $x$  and  $y$  represent the sine values of the angles between the center points of the ground truth and predicted boxes. At the same time,  $\rho$  is the ratio of the distance between the center points of the ground truth and predicted boxes relative to the width and height of their minimum enclosing rectangle squared. Here,  $e$  denotes the Euler's number.

$$\Omega = \sum_{t=w,h} (1 - e^{-\omega t})^\theta \quad (4)$$

The above equation computes the  $\Omega$  shape loss, where  $w$  and  $h$  represent the width and height of the predicted box, respectively, and  $\theta$  denotes the attention coefficient in the shape calculation formula. Finally, the SIOU loss consists of the three components above, as depicted by the following formula:

$$L_{box} = 1 - IoU + \frac{\Omega + \Delta}{2} \quad (5)$$



## 4. Experimental results and analysis

### 4.1. Enhancement and partitioning of experimental datasets

On high-speed cigarette production lines, the probability of encountering cigarettes with visual defects is approximately 1%. The manual screening process to obtain a substantial amount of defective cigarette data is labor-intensive. We gathered 900 valid images of appearance defect cigarettes through meticulous manual selection. Due to the limited availability of the appearance defect dataset, we employed data augmentation techniques, including flipping, brightness adjustment and Gaussian noise addition, to expand the dataset. Consequently, the augmented dataset now comprises 6200 images. It is important to note that since the images were captured in pairs, the actual research dataset consists of 12,400 cigarette images. To ensure the model's effectiveness, we divide the dataset into training, validation and testing sets in a 6:2:2 ratio, as detailed in Table 1.

**Table 1.** Dataset partition.

Defect type	Training set	Validation set	Testing set
Misplacement	480	160	160
Stain	840	280	280
Scratched	1680	560	560
Wrinkle	1800	600	600
Unpacked	1476	492	492
No filter	876	292	292
Bend	420	140	140
Total	7572	2524	2524

### 4.2. Experiment parameter setting

The model was trained and tested on a Windows 10 system running PyTorch 1.12.1, using the following hardware specifications: an AMD R5600 processor @ 3.50 GHz, 32 GB of memory and an NVIDIA GeForce RTX3060 graphics card with 12 GB of VRAM. The software was CUDA 11.6, Torch vision 0.13.1 and Python 3.7. The integrated development environments are PyCharm and Anaconda.

**Table 2.** Training parameter.

Parameter	Value
Image size	640 × 640
Batch size	64
Epoch	300
Learning rate	0.01
Optimizer	SGD
Weight decay	0.0005

The initial learning rate was 0.01 during training, and a cosine annealing strategy was employed to reduce the learning rate. Additionally, the neural network parameters were optimized using the

stochastic gradient descent (SGD) method with a momentum value of 0.937 and a weight decay score of 0.0005. The training process consisted of 300 epochs, with a batch size of 64 images. The input image resolution was uniformly adjusted to  $640 \times 640$ . The adjusted training parameters are summarized in Table 2.

#### 4.3. Evaluation indicators

The experimental evaluation encompasses two aspects: performance evaluation and complexity evaluation. Performance evaluation metrics for the model include accuracy, recall, mAP@0.5 and mAP@0.5-0.95. Complexity evaluation metrics for the model consist of the size of model parameters, floating point operations (FLOPs) and frames per second (FPS), which assess the computational efficiency and image processing speed of the model.

Precision measures the proportion of correctly predicted positive samples out of the total positive samples, assessing the model's classification ability. Conversely, recall measures the proportion of correctly predicted positive samples out of the whole of positive samples. AP is the integral of precision and recall, and mAP represents the average AP, reflecting the model's overall performance for object detection and classification. The calculation formulas for these metrics are shown in Eqs (6)–(9).

$$Precision = \frac{TP}{TP + FP} \quad (6)$$

$$Recall = \frac{TP}{TP + FN} \quad (7)$$

where,  $TP$  represents the number of true positive samples correctly detected,  $FP$  represents the number of false positive samples incorrectly detected, and  $FN$  represents the number of false negative samples incorrectly not detected.

$$AP = \int_0^1 P(R) dR \quad (8)$$

$$mAP = \frac{\sum_{i=1}^n AP_i}{n} \quad (9)$$

where,  $n$  is the number of categories.

$$FLOPs(Conv) = (2 \times C_{in} \times K^2 - 1) \times W_{out} \times H_{out} \times C_{out} \quad (10)$$

$$FLOPs(Liner) = (2 \times C_{in} - 1) \times C_{out} \quad (11)$$

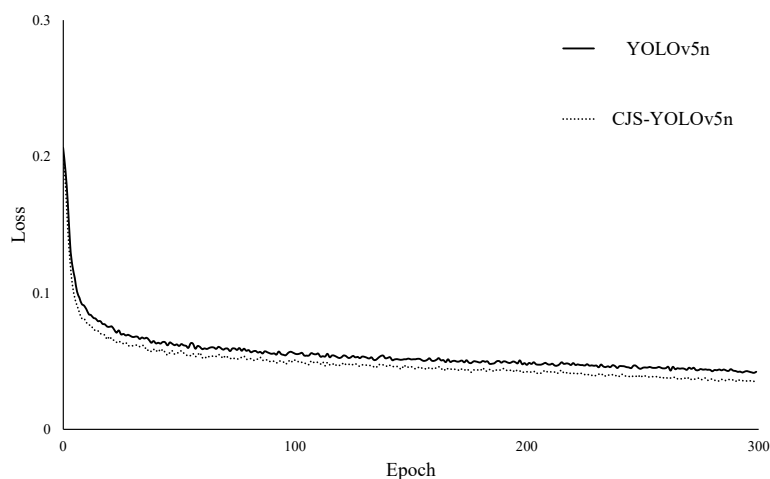
Model size refers to the amount of memory required to store the model. FLOPs, on the other hand, measure the complexity of the model by quantifying the total number of multiplication and addition operations performed during model execution. A lower FLOPs value indicates lower computational requirements for model inference, resulting in faster model computation speed. Here,  $C_{in}$  represents the number of input channels,  $C_{out}$  represents the number of output channels,  $K$

represents the convolutional kernel size and  $W_{out}$  and  $H_{out}$  represent the width and height of the output feature map, respectively.

#### 4.4. Training process analysis

Figure 8 presents the training loss curves for the CJS-YOLOv5n method and YOLOv5n on the cigarette appearance defect dataset. The graph illustrates the overall loss values during the training process. In the initial 30 epochs, the model experiences a rapid decline in loss. Subsequently, from epoch 50 to 200, a gradual decrease in loss is observed. Between epochs 200 and 300, the loss values stabilize and approach convergence. Thus, 300 epochs are determined as the appropriate training iteration count for the model.

Furthermore, the dashed line represents the loss curve of the CJS-YOLOv5n, while the solid line represents the loss curve of YOLOv5n. It can be observed from the graph that, under the same conditions of rapid loss reduction and convergence, the improved model exhibits a lower final convergence value compared to the original model. Additionally, our improved model shows more minor fluctuations in the loss curve during training. Consequently, our improved model demonstrates superior performance during the training process.



**Figure 7.** Training loss curve.

#### 4.5. Ablation experiment

This section verifies the effectiveness of different optimization modules through ablation experiments. Several improved models are constructed sequentially, adding the C2F module, Jump Concat and SIOU localization loss function to the baseline model YOLOv5n. The results are compared using the same test data, and the gains in model performance due to the added modules are presented in Table 3.

Table 3 shows the gain of model performance after adding each module. Due to the small depth and width of the original YOLOv5n network, the ability to fit complex features is poor. In order to enhance the feature extraction ability of the network, this paper replaces the C3 convolution module with the C2F convolution module, which increases mAP@0.5 by 0.6%. The rate increased by 0.6%.

Later, in order to prevent the feature fusion from covering up the fine texture features, this paper adds a Jump Concat to the P4 feature layer, which increases mAP@0.5 by 1% and the recall rate by 1%. Finally, in order to enhance the positioning accuracy and avoid false detection, this paper replaces the CIoU loss function with the SIoU loss function, which increases mAP@0.5 by 1% and the recall rate by 1.1%. After combining the three modules, the improved model achieves the best performance. The increased computational load of the improved module in this paper does not affect the performance of the model running on the GPU, but the increased parallel computing of the C2F convolution module makes the running speed of the model on the CPU drop by nearly 10%.

**Table 3.** Ablation experiment result.

C2F	Jump Concat	SIoU	Precision (%)	Recall (%)	mAP@0.5 (%)	mAP@0.5:0.95 (%)	FLOPs (G)	FPS (GPU)	FPS (CPU)
			94.8	93	94.2	56	4.2	556	43
✓			95.1	93.6	94.8	56.4	5.6	556	40
	✓		94.9	94	95.2	56.5	4.2	556	43
		✓	94.2	94.1	95.2	56.9	4.2	556	43
✓	✓		95.5	93.9	95.7	57.7	5.6	556	40
✓		✓	<b>95.5</b>	94.2	95.1	57.2	5.6	556	40
	✓	✓	94.7	95	95.5	56.8	4.2	556	43
✓	✓	✓	95.3	<b>95.3</b>	<b>95.9</b>	<b>57.9</b>	5.6	556	40

#### 4.6. Comparative experiments

The proposed method in this study was trained using the same parameters as other advanced lightweight methods. The experimental results are compared in Table 4.

**Table 4.** Comparison of different detection models.

Model	Precision (%)	Recall (%)	mAP@0.5 (%)	mAP@0.5:0.95 (%)	Parameters (M)	FPS
YOLOv3tiny	82.3	93.7	92.9	54.3	8.6	500
YOLOv4tiny	88.4	93.2	92.8	54.0	5.9	500
YOLOv5n	94.8	93.0	94.2	56.0	1.7	556
YOLOv7tiny	91.5	87.0	91.6	51.3	6.2	278
YOLOv8n	<b>95.7</b>	93.3	95.4	57.2	3.0	286
SSD-improved [16]	87.4	88.7	90.3	50.1	26	84
YOLOv4-improved [18]	95.2	92.8	94.8	55.7	66	161
CJS-YOLOv5n	95.3	<b>95.3</b>	<b>95.9</b>	<b>57.9</b>	2.3	556

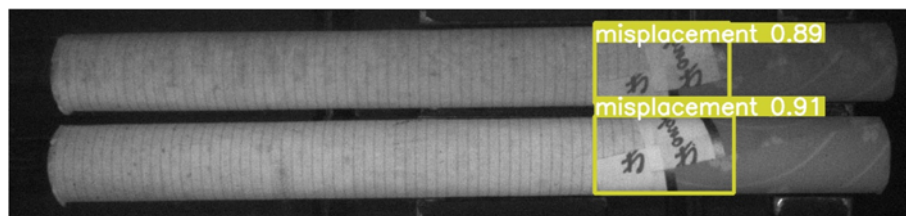
Table 4 shows the performance comparison with other target detection models. The model selected in this paper is not optimal in every performance indicator, but it is faster than the better

performance models like YOLOv8n, and detection accuracy is higher than models like YOLOv3tiny that are close in speed. Since the number of model parameters selected in this paper is the least among all models, it is beneficial to reduce the deployment cost of practical applications and achieve a balance between detection speed and accuracy. While maintaining the detection speed, the improved CJS-YOLOv5n in this paper has improved mAP@0.5 by 1.7% and the recall rate by 2.3% compared with the original model. YOLOv8 is a state-of-the-art object detection model with high accuracy and speed in many applications. Compared with the most advanced YOLOv8n in average detection accuracy, recall rate and detection speed achieved an all-around lead, which can meet most of the current detection needs. The experimental results show that the model proposed in this paper is a good detection method for cigarette appearance defects, which can meet the needs of cigarette appearance defect detection on the production line.

Our ablation experiments and comparative experiments demonstrate the effectiveness of our improvements. Among them, the C2F module achieves better feature extraction at the cost of a small amount of calculation, and various performance indicators have different improvements. Jump Concat reduces the degree of occlusion of subtle features in feature fusion, effectively improving the recall rate. The SIOU loss function improves the localization accuracy of anchor boxes by optimizing the anchor box loss calculation, thus increasing the confidence and recall of predicted anchor boxes. Comparing the latest research results of YOLOv4-improved in detecting cigarette appearance defects and the current state-of-the-art object detection model YOLOv8, the improved model CJS-YOLOv5n has different degrees of increase in detection performance and detection speed.

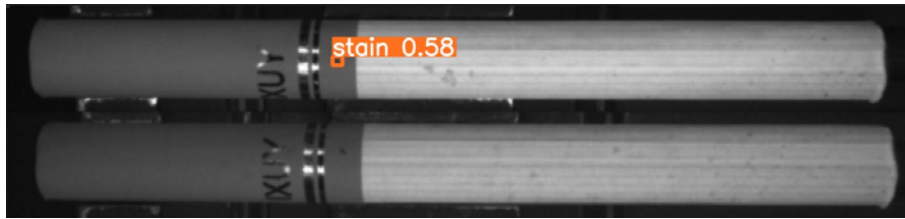
#### 4.7. Detection results comparison

Figure 9(a)–(h) shows the detection results of YOLO-v5n for seven kinds of defects, and Figure 10(a)–(h) shows the detection results of CJS-YOLOv5n for seven kinds of defects. The detection results prove the improvement of this paper. Compared with the original model, the latter model has better performance in seven kinds of defect detection. Comparing Figure 9 and (b), (d) and (h) in Figure 10, it can be found that the improved model can better identify the subtle wrinkle and stain defects that were not detected by the original network, and the positioning anchor box of the fold defect is more complete. In complex scenarios with multiple defects, the improved model also avoids the false detection of the original model. Comparing the detection results of other defects, it can be seen that the improved network has different confidence improvements compared with the original network in the detection of different defects.

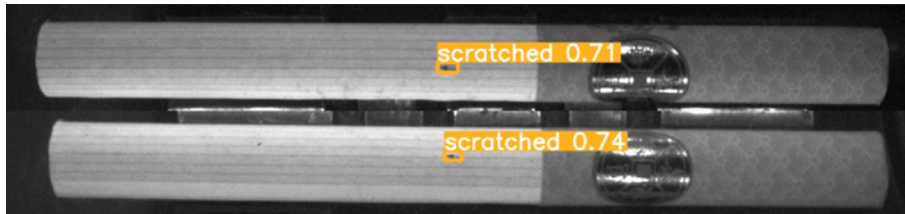


(a) Misplacement

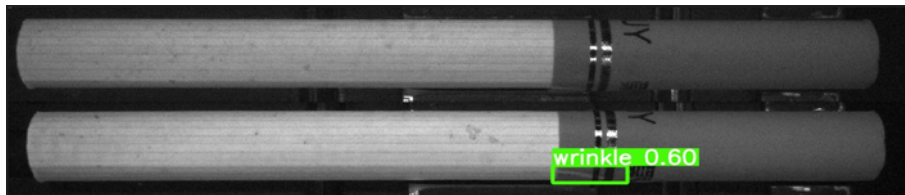
*Continued on next page*



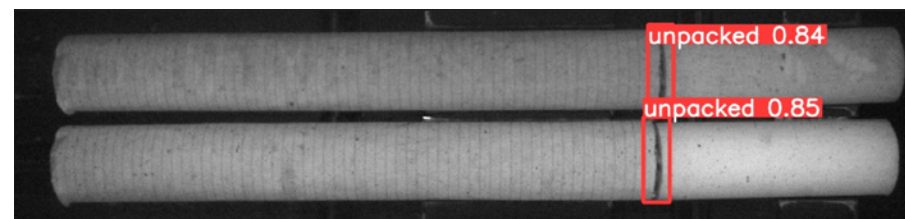
(b) stain



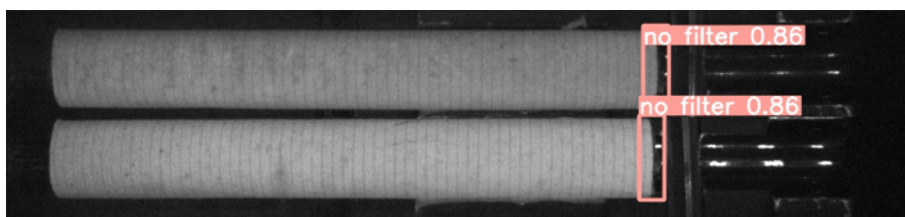
(c) scratched



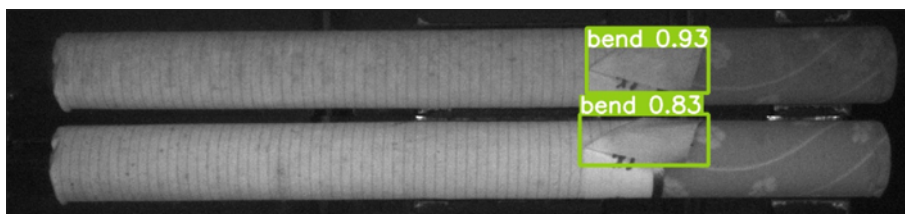
(d) wrinkle



(e) unpacked

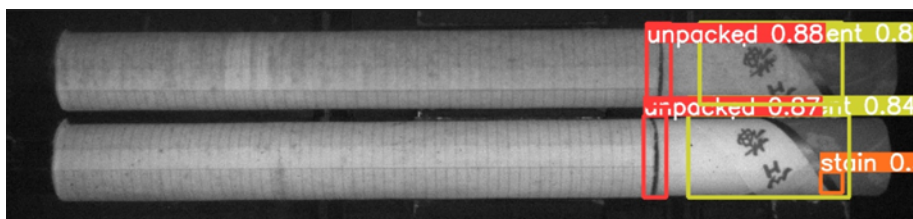


(f) no filter

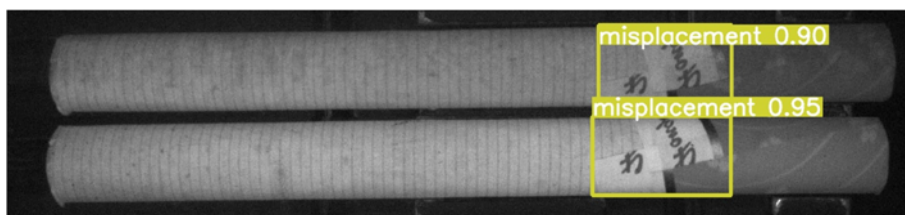


(g) bend

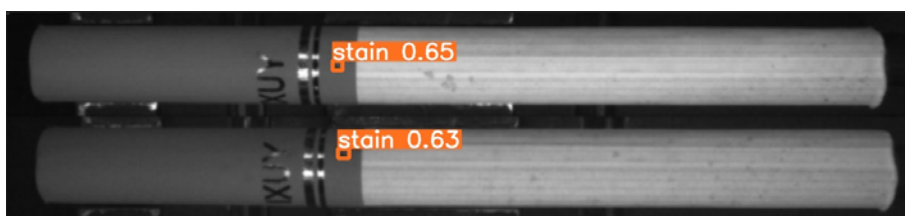
*Continued on next page*



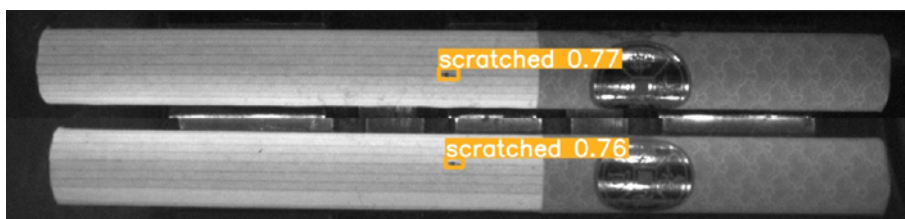
(h) Multiple appearance defects

**Figure 8.** YOLOv5n detection result.

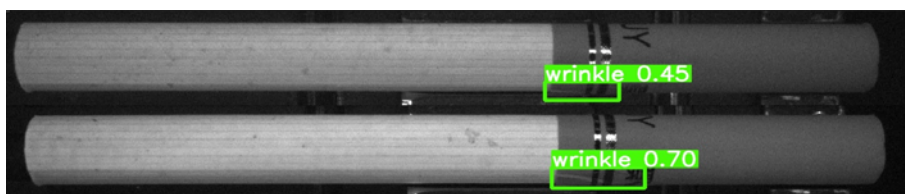
(a) misplacement



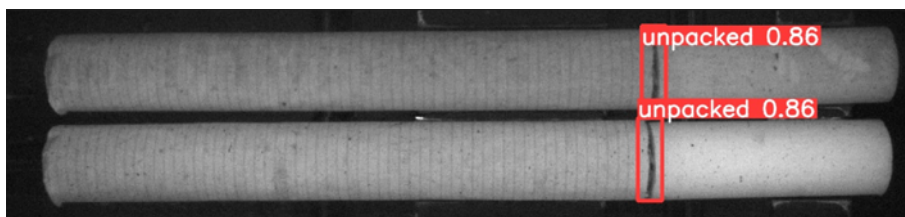
(b) stain



(c) scratched

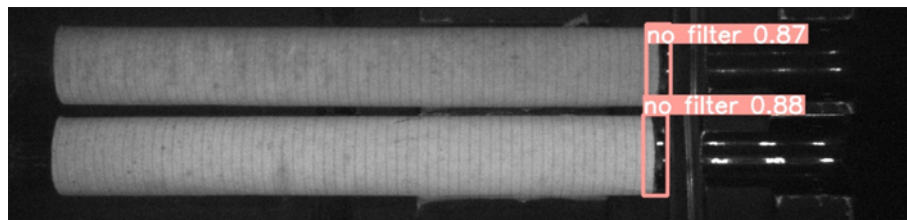


(d) wrinkle

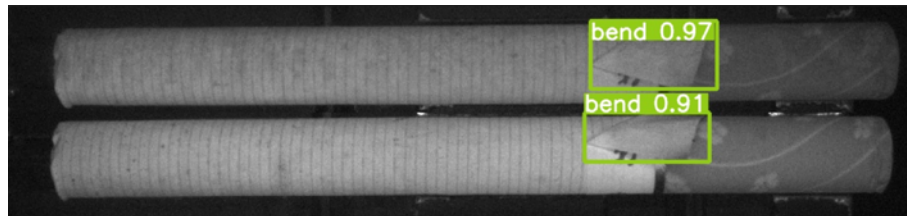


(e) unpacked

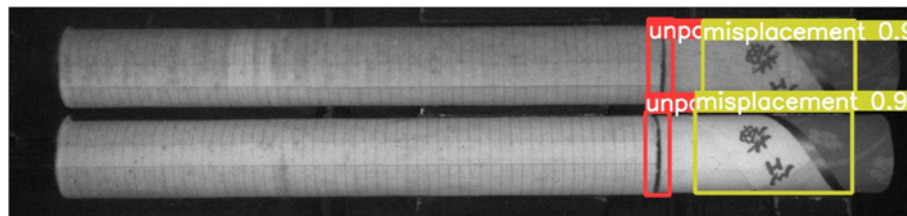
*Continued on next page*



(f) no filter



(g) bend



(h) Multiple appearance defects

**Figure 9.** CJS-YOLOv5n detection result.

## 5. Conclusions

According to the fact that the actual production speed of cigarette appearance defects is fast, the improved YOLOv5n network is used to detect cigarette appearance defects, and good detection results and detection speed can be achieved in the case of insufficient data sets. In this paper, the C2F module with stronger feature extraction ability than the C3 convolution module is selected as the network basic convolution module, and the P4 feature layer Jump Concat is added to further strengthen the feature fusion, prevent information coverage and use the SIoU positioning loss function to help the model improve. The positioning accuracy and convergence speed in the experimental results prove the effectiveness of the algorithm in the task of cigarette appearance defect detection. This method can achieve better detection speed and recall rate, and has certain robustness. The model can effectively improve the recall rate and detection accuracy without affecting the detection speed.

Our model in this paper still has shortcomings. For example, in terms of detection effect, it is still not as good as YOLOv8n with a larger network. Although the improvement in this paper can effectively increase the recall rate and help control the quality of cigarette sales, it is limited by the depth and width of the network. The ability to fit complex defects still needs to be improved. In the future, we will focus on how to improve the accuracy of network detection while maintaining the detection speed and reducing the deployment cost, such as adding a lightweight attention mechanism CBAM (Convolutional block attention module) [31], a detection head based on Anchor-Free [32], further expand the dataset, and use lightweight convolution. In addition, we will use higher-order moment time



series analysis in the Caputo sense [33] and fractional differential equations [34] to improve post-processing. We hope we can achieve better performance in cigarette appearance defect detection.

### Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

### Acknowledgements

Thank you to Yunnan Branch of China Tobacco Industry Company Limited for providing the cigarette defect dataset.

This research was funded by the Yunnan Provincial Department of Science and Technology-Yunnan University Joint Special Project for Double-Class Construction (Grant No. 202201BF070001-005), the Natural Science Foundation of China (Grant No. 62061049, 12263008), the Key R&D Projects of Yunnan Province (Grant No. 202202AD080004), the Application and Foundation Project of the Yunnan Province (Grant No. 202001BB050032) and the Postgraduate Practice and Innovation Project of Yunnan University (Grant no. 22221382).

### Conflict of interest

The authors declare there is no conflict of interest.

### References

1. R. Girshick, J. Donahue, T. Darrell, J. Malik, Rich feature hierarchies for accurate object detection and semantic segmentation, in *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition*, (2014), 580–587. <https://doi.org/10.1109/CVPR.2014.81>
2. R. Girshick, Fast R-CNN, in *Proceedings of the International Conference on Computer Vision*, (2015), 1440–1448. <https://doi.org/10.1109/ICCV.2015.169>
3. S. Ren, K. He, R. Girshick, J. Sun, Faster R-CNN: Towards real-time object detection with region proposal networks, *IEEE Trans. Pattern Anal. Mach. Intell.*, **39** (2017), 1137–1149. <https://doi.org/10.1109/TPAMI.2016.2577031>
4. J. Redmon, S. Divvala, R. Girshick, A. Farhadi, You only look once: Unified, real-time object detection, in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (2016), 779–788. <https://doi.org/10.1109/CVPR.2016.91>
5. J. Redmon, A. Farhadi, YOLO9000: Better, faster, stronger, in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (2017), 6517–6525. <https://doi.org/10.1109/CVPR.2017.690>
6. J. Redmon, A. Farhadi, YOLOv3: An incremental improvement, preprint, arXiv:1804.02767.
7. A. Bochkovskiy, C. Y. Wang, H. Liao, YOLOv4: Optimal speed and accuracy of object detection, preprint, arXiv:2004.10934
8. W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Y. Fu, et al., SSD: single shot multibox detector, in *Proceedings of the European Conference on Computer Vision*, **9905** (2016), 21–37. [https://doi.org/10.1007/978-3-319-46448-0\\_2](https://doi.org/10.1007/978-3-319-46448-0_2)

9. J. Liu, X. Zhu, X. Zhou, S. Qian, J. Yu, Defect detection for metal base of TO-Can packaged laser diode based on improved YOLO algorithm, *Electronics*, **11** (2022), 1561. <https://doi.org/10.3390/electronics11101561>
10. C. S. Han, T. C. Chun, SMD LED chips defect detection using a YOLOv3-dense model, *Adv. Eng. Inf.*, **47** (2021), 101255. <https://doi.org/10.1016/j.aei.2021.101255>
11. B. Hu, J. H. Wang, Detection of PCB surface defects with improved faster-RCNN and feature pyramid network, *IEEE Access*, **8** (2020), 108335–108345. <https://doi.org/10.1109/ACCESS.2020.3001349>
12. Y. Li, J. Xu, Electronic product surface defect detection based on a MSSD network, in *2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*, (2020), 773–777. <https://doi.org/10.1109/ITNEC48623.2020.9084756>
13. L. M. Duan, Y. Ke, R. Lang, Research on automatic recognition of casting defects based on deep learning, *IEEE Access*, **9** (2020), 12209–12216. <https://doi.org/10.1109/ACCESS.2020.3048432>
14. X. P. Kou, S. J. Liu, K. Q. Cheng, Y. Qian, Development of a YOLO-V3-based model for detecting defects on steel strip surface, *Measurement*, **182** (2021), 109454. <https://doi.org/10.1016/j.measurement.2021.109454>
15. G. W. Yuan, J. C. Liu, H. Y. Liu, R. Qu, H. Zhou, Classification of cigarette appearance defects based on ResNeSt, *J. Yunnan Univ.: Nat. Sci. Ed.*, **44** (2022), 464–470. <https://doi.org/10.7540/j.ynu.20210257>
16. R. Qu, G. W. Yuan, J. C. Liu, H. Zhou, Detection of cigarette appearance defects based on improved SSD model, in *Proceedings of the 2021 5th International Conference on Electronic Information Technology and Computer Engineering (EITCE '21)*, (2021), 1148–1153. <https://doi.org/10.1145/3501409.3501612>
17. H. Y. Liu, G. W. Yuan, Cigarette appearance defect detection method based on improved YOLOv5s, *Comput. Technol. Dev.*, **32** (2022), 161–167. <https://doi.org/10.3969/j.issn.1673-629X.2022.08.026>
18. G. W. Yuan, J. C. Liu, H. Y. Liu, Y. H. Ma, H. Wu, H. Zhou, Detection of cigarette appearance defects based on improved YOLOv4, *Electr. Res. Arch.*, **31** (2023), 1344–1364. <https://doi.org/10.3934/era.2023069>
19. H. Y. Liu, G. W. Yuan, L. Yang, K. X. Liu, H. Zhou, An appearance defect detection method for cigarettes based on C-CenterNet, *Electronics*, **11** (2022), 2182. <https://doi.org/10.3390/electronics11142182>
20. D. Y. Xu, G. D. Deng, C. X. Liu, F. Y. Shu, J. F. Yan, Q. Zhou, Particle size analysis of cigarette-puncturing slivers, *Acta Tabacaria Sinica*, **27** (2021), 37–42. <https://doi.org/10.16472/j.chinatobacco.2020.148>
21. Y. W. Li, J. J. Qiao, S. J. Ma, Z. Q. Wu, Q. H. Wu, Improvement of filter plug assembling quality in cigarette making, *Tob. Sci. Technol.*, **45** (2012), 24–27. <https://doi.org/10.3969/j.issn.1002-0861.2012.10.006>
22. G. Jocher, A. Stoken, J. Borovec, X. Tao, Y. Kwon, K. Michael, et al., *Ultralytics/YOLOv5: V6.0—YOLOv5n 'Nano' Models, Roboflow Integration, TensorFlow Export, OpenCV DNN Support*, 2021. Available from: <https://zenodo.org/record/5563715>.
23. Z. Zheng, P. Wang, W. Liu, J. Li, R. Ye, D. Ren, Distance-IoU loss: Faster and better learning for bounding box regression, in *Proceedings of the Thirty-Fourth AAAI Conference on Artificial Intelligence (AAAI-20)*, **34** (2020), 12993–13000. <https://doi.org/10.1609/aaai.v34i07.6999>

24. J. Terven, D. C. Esparza, A comprehensive review of YOLO: From YOLOv1 to YOLOv8 and beyond, preprint, arXiv:2304.00501.
25. H. Park, Y. Yoo, G. Seo, D. Han, S. Yun, N. Kwak, C3: Concentrated-comprehensive convolution and its application to semantic segmentation, preprint, arXiv:1812.04920.
26. Z. Gevorgyan, SIOU loss: More powerful learning for bounding box regression, preprint, arXiv:2205.12740.
27. T. Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, S. Belongie, Feature pyramid networks for object detection, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (2017), 2117–2125. <https://doi.org/10.1109/CVPR.2017.106>
28. S. Liu, L. Qi, H. Qin, J. Shi, J. Jia, Path aggregation network for instance segmentation, in *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, (2018), 8759–8768. <https://doi.org/10.1109/cvpr.2018.00913>
29. L. T. Jiao, B. Hong, An optimized YOLO method for object detection, in *2020 16th International Conference on Computational Intelligence and Security (CIS)*, (2020), 30–34. <https://doi.org/10.1109/CIS52066.2020.00015>
30. H. Rezatofghi, T. Nathan, J. Y. Gwak, S. Amir, R. Ian, S. Silvio, Generalized intersection over union: A metric and a loss for bounding box regression, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (2019), 658–666. <https://doi.org/10.1109/cvpr.2019.00075>
31. S. Woo, J. Park, J. Y. Lee, I. S. Kweon, CBAM: Convolutional block attention module, in *Proceedings of the European Conference on Computer Vision (ECCV)*, **11211** (2018), 3–19. [https://doi.org/10.1007/978-3-030-01234-2\\_1](https://doi.org/10.1007/978-3-030-01234-2_1)
32. C. C. Zhu, Y. H. He, S. Marios, Feature selective anchor-free module for single-shot object detection, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, (2019), 840–849. <https://doi.org/10.1109/cvpr.2019.00093>
33. C. Hao, T. Jin, F. L. Tan, J. R. Gao, Z. X. Ma, J. Cao, The analysis of time-varying high-order moment of wind power time series, *Energy Rep.*, **9** (2023), 3154–3159. <https://doi.org/10.1016/j.egy.2023.02.010>
34. T. Jin, F. Li, H. Peng, B. Li, D. Jiang, Uncertain barrier swaption pricing problem based on the fractional differential equation in Caputo sense, *Soft Comput.*, **27** (2023), 11587–11602. <https://doi.org/10.1007/s00500-023-08153-5>



AIMS Press

©2023 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)