




Article

SCS-YOLO: A Defect Detection Model for Cigarette Appearance

Yingchao Ding ¹, Hao Zhou ¹ , Hao Wu ¹, Chenrui Ma ²  and Guowu Yuan ^{1,*} 

¹ School of Information Science and Engineering, Yunnan University, Kunming 650504, China; dingyingchao@stu.ynu.edu.cn (Y.D.); zhouhao@ynu.edu.cn (H.Z.); haowu_sise@ynu.edu.cn (H.W.)

² School of Computer Science and Engineering, Central South University, Changsha 410083, China; 8208210320@csu.edu.cn

* Correspondence: gwyuan@ynu.edu.cn

Abstract: Appearance defects significantly impact cigarette quality. However, in the current high-speed production lines, manual inspection and traditional methods are unable to satisfy the actual demands of inspection. Therefore, a real-time and high-precision defect detection model for cigarette appearance, SCS-YOLO, is presented. The model integrates space-to-depth convolution (SPD-Conv), a convolutional block attention module (CBAM), and a self-calibrated convolutional module (SC-Conv). SPD-Conv replaces the pooling structure to enhance the granularity of feature information. CBAM improves the ability to pay attention to defect locations. Improved self-calibrated convolution broadens the network's receptive field and feature fusion capability. Additionally, Complete IoU loss (CIoU) is replaced with Efficient IoU Loss (EIoU) to enhance model localization and mitigate sample imbalance. The experimental results show that the accuracy of SCS-YOLO is 95.5% and the mAP (mean average precision) value is 95.2%. Compared with the original model, the accuracy and mAP value of the SCS-YOLO model are improved by 4.0%. Furthermore, the model achieves a detection speed of 216 FPS, meeting cigarette production lines' accuracy and speed demands. Our research will positively impact the real-time detection of appearance defects in cigarette production lines.

Keywords: cigarette; defect detection; object detection; YOLOv7-tiny; attention mechanism

MSC: 68T45



Citation: Ding, Y.; Zhou, H.; Wu, H.; Ma, C.; Yuan, G. SCS-YOLO: A Defect Detection Model for Cigarette Appearance. *Electronics* **2024**, *13*, 3761. <https://doi.org/10.3390/electronics13183761>

Academic Editor: Jenhui Chen

Received: 15 August 2024

Revised: 11 September 2024

Accepted: 20 September 2024

Published: 22 September 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

China is a major tobacco-producing country. Although tobacco poses certain hazards to human health, the tobacco industry contributes significantly to China's tax revenue, raises farmers' incomes, and remains crucial for maintaining the country's economic stability. Cigarettes are the most important products of the tobacco industry. On large-scale high-speed assembly lines, various appearance defects may occur for multiple reasons; these defects include anomalies that the literature labels "Dotted", "Folds", "Untooth", and "Unfilter" [1]. These appearance defects seriously affect the quality of cigarettes and damage the brand. Therefore, tobacco companies must strictly control quality and avoid selling cigarettes with appearance defects.

Early quality control in cigarette factories relied on manual inspection. However, on a high-speed production line, cigarettes are produced at a rate of 150–200 per second, with each line producing about 15 million cigarettes per day [2]. With such a high production rate, manual inspection is impractical, creating an urgent need for automated cigarette appearance inspection technology to replace manual methods.

With the development of machine vision, manual inspection of product quality has been gradually replaced by automatic inspection, significantly improving the efficiency of industrial inspections, such as printed circuit board defect detection, packaging seal defect detection, welding defect detection, photovoltaic component defect detection, and steel surface defect detection. Product surface defect detection falls under the category of

object detection. At present, the most widely used object detection algorithms are Faster R-CNN [3–5], SSD [6], YOLO [7–10], etc. Practitioners in various industries have improved these methods to adapt to object detection tasks.

Domestic and international research on cigarette appearance defect detection is relatively limited. However, significant advancements have been made in other appearance defect and defective product detection techniques, which can be leveraged for cigarette appearance detection tasks. Zhang et al. [11] used YOLOv3 to detect surface defects on aluminum profiles. Cheng et al. [12] proposed an algorithm to improve YOLOv3 for metal surface defect detection by generating a large-scale feature layer to extract more small target features. Wei et al. [13] combined SSD with the MobileNet algorithm and then designed an SSD-MobileNet model for surface defect detection on manufactured boards. Peng et al. [14] used the Res-Net101 backbone network based on Faster R-CNN to detect surface defects on particleboard. Luo et al. [15] introduced the ResNet18 structure into YOLOv3, significantly improving the accuracy and detection speed of the model for UAV power grid inspection. Hu et al. [16] put forward a detection method to detect pipeline weld surface defects based on multi-feature extraction and binary tree support vector machine (BT-SVM) classification. Lu et al. [17] incorporated a multi-scale detection network based on FPN and GA-RPN structure into Faster R-CNN, achieving a mAP of 92.5% in solar cell defect detection. Hu et al. [18] added a pyramid pooling module and a spatial and coordinate attention module to the YOLOv5 network to improve the detection of small-size defects, achieving an average accuracy of 97.6% in detecting surface defects on rebar. Wang et al. [19] designed a defect detection method based on a watershed algorithm and two-pathway convolutional neural network and applied it to corn seed defect detection. The average accuracy of the experiment reached 95.63%. Wan et al. [20] improved the detection ability of small-size defects by adding a small object detection layer in the YOLOv5s network, effectively solving problems caused by small-size defects and insufficient feature information in tile surface defect detection. Zhang et al. [21] embedded the SENet module into the Unet network to detect appearance defects of parts, achieving a detection accuracy of 92.98%. Qi et al. [22] combined the YOLOv7-tiny network with a weighted Bidirectional Feature Pyramid Network (BiFPN) for steel surface defects, effectively improving the efficiency of small object detection with an average accuracy of 94.6%. Jing et al. [23] proposed a Mobile-Unet network to realize end-to-end fabric defect detection and introduced depth-wise separable convolution, reducing model complexity and achieving an accuracy rate of 92%. Li et al. [24] designed a lightweight convolutional neural network, WearNet, to realize automatic surface scratch detection, achieving an excellent classification accuracy of 94.16% with smaller model size and faster detection speed.

In recent years, some scholars have researched cigarette appearance defect detection. Qu et al. [25] introduced a method for detecting cigarette appearance defects utilizing an improved SSD model, using the ResNet50 network for feature extraction and adding pyramid convolution to achieve a mAP of 94.54%. However, the ResNet50 model only increases the network depth by simply stacking the residual blocks, resulting in a relatively simple network structure. This simplicity leads to insufficient feature extraction capability and generates many redundant features, as a result, the model's recall is only 69.83%. The deeper network structure also reduces the inference speed, with the detection speed being only 66 FPS. Yuan et al. [26] proposed a classification method for cigarette appearance defects based on ResNeSt, utilizing a transfer learning method to train ResNeSt to address the problem of an insufficient number of samples. However, the ResNeSt network uses the Split-Attention Block to slice the feature channels, reducing the correlation, complementarity, and information interaction ability between the channels, thus weakening the feature fusion ability of the model. Yuan et al. [27] proposed a defect detection method for cigarette appearance based on improved YOLOv4; by introducing the channel attention mechanism and replacing the spatial pyramid pooling structure (SPP) with atrous spatial pyramid pooling structure (ASPP), a mAP of 91.77% was achieved. However, introducing these modules increased the computational complexity of the model, and the detection

speed was lower than that of the original YOLOv4 model, which was only 53 FPS. Liu et al. [1] proposed a defect detection method for cigarette appearance based on improved YOLOv5s. The original model was improved by data augmentation, the introduction of the channel attention mechanism, and optimization of the activation function and loss function. However, since YOLOv5s is a lightweight model in the YOLOv5 series, its feature extraction capability needs to be improved, and all improvements focus on improving the detection speed rather than the algorithm's capacity for learning features. As a result, its recall is only 86.8%. Liu et al. [2] proposed a cigarette defect detection method based on C-CenterNet, which improves the detection ability of the algorithm for small target defects, by introducing the CBAM attention mechanism, replacing ordinary convolution with deformable convolution, and adopting the feature pyramid network (FPN). However, since the FPN uses a top-down approach to feature fusion, it cannot fully utilize low-level feature information, therefore, the recall is only 85.96%, and the detection speed is reduced to 112 FPS, which cannot meet the real-time requirements.

The above methods are significant for researching cigarette appearance defect detection. However, most of them need improvement in some areas, such as low detection accuracy due to insufficient feature extraction and poor feature fusion capabilities and slow detection speed caused by increased model complexity or the introduction of complex calculations. In addition, the dataset that has been obtained still needs help with small samples and unbalanced samples. Because YOLOv7 [28] has shown high accuracy and fast detection speed in many object detection tasks, in order to meet the requirements of high-speed production lines of cigarettes, this paper uses YOLOv7 as the primary network and proposes a new model of cigarette appearance defect detection by introducing space-to-depth convolution, a convolutional block attention module, and a self-calibrated convolution module; the resulting model is called SCS-YOLO. The main contributions of this paper are as follows:

- (1) The space-to-depth convolution is introduced into the backbone network part of the model, which achieves feature dimension reduction and reduces the loss of small target features. Its addition method is analyzed and experimented with in detail. We found that although the space-to-depth convolution can retain the information of small targets to the greatest extent, it damages the integrity of the information about large targets. Therefore, a pooling layer and two space-to-depth convolutional layers are used in the backbone network to increase the fine-grained information in the network while ensuring the integrity of the large targets' information.
- (2) To enhance the model's ability to pay attention to target information, the convolutional block attention modules are introduced after the two C5 modules in the backbone network, which improves the model's ability to extract small targets and distinguish similar defects.
- (3) The dual self-calibrated convolutional module (D-SCConv) is proposed and applied to the neck network; it enhances the receptive field of each region in the feature map and calibrates the high-level semantic information, thereby helping the model to generate higher-quality features.
- (4) The EIou [29] loss function is used to replace the CIou [30] loss function in the original model, which accelerates the convergence speed of the model, improves the positioning ability of the model, and reduces the impact of sample imbalance.

The experimental results show that the SCS-YOLO model has improved the detection of various cigarette appearance defects and meets the requirements of industrial production lines.

2. Introduction to Cigarette Appearance Defects

Cigarettes are the most important product of the tobacco industry. However, tobacco stems in the raw materials, deviations in machine processing, and machine failures may cause various appearance defects on a high-speed production line.

This paper constructs a dataset of cigarette appearance defects based on the actual needs of the Yunnan Branch of China Tobacco Industry Company Limited. The images were captured using a high-speed industrial camera installed on the production lines.

A cigarette consists of the cigarette rod and the filter tip. The cigarette rod is longer and holds the tobacco, while the filter tip is shorter and holds the filter cotton. A standard cigarette is 84 mm in length and 7.8 mm in diameter, resulting in an aspect ratio of about 10:1. Cigarette appearance defects can be categorized into four groups based on their manifestations, designated herein as “Dotted”, “Folds”, “Unfilter”, and “Untooth”. “Dotted” defects consist of minor wrapper stains or are caused by thorny stems in the tobacco that puncture the cigarette rod. “Folds” defects are caused by improper squeezing and twisting of the cigarette paper by production machines, resulting in cigarette paper folds. “Untooth” refers to the misalignment of cigarette or filter paper during production, usually caused by loosening of the packing machinery. “Unfilter” refers to the phenomenon where the filter is detached or cannot be wrapped correctly due to depletion of the filter paper, machine failure, or insufficient adhesive properties of the latex. Furthermore, “Normal” cigarettes are defined as those without appearance defects. Images of appearance defects are shown in Figure 1.

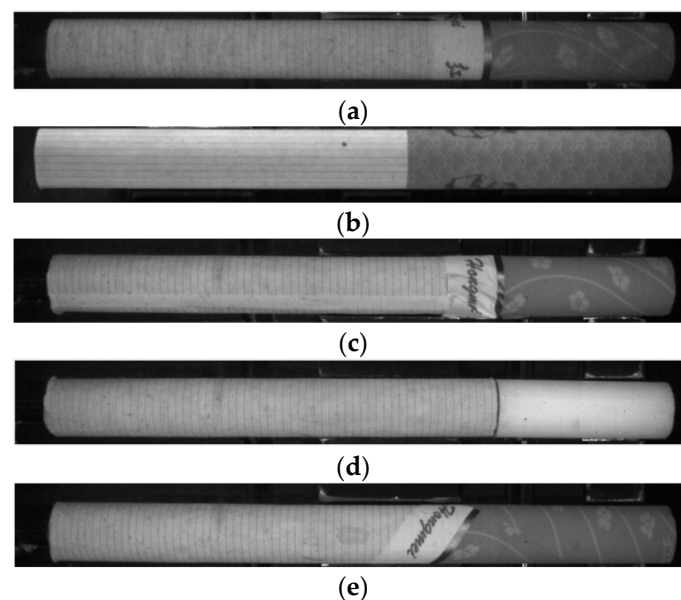


Figure 1. Cigarette images with appearance defects: (a) Normal; (b) Dotted; (c) Folds; (d) Unfilter; (e) Untooth.

3. Proposed Method

In July 2022, the original team that developed YOLOv4 proposed YOLOv7. YOLOv7 achieves a twofold improvement in detection speed and detection accuracy, enabling it to be applied to real-time object detection tasks. The YOLOv7 network consists of four components: input, backbone, neck, and head. The input module preprocesses the input image, zooming processing, and other operations to ensure it meets the network’s input requirements. Next, the processed image is fed into the backbone network, which extracts features from the images. The neck module then performs feature fusion on the extracted features, forming three different scales: large, medium, and small. Finally, these fused features are sent to the detection head, where the detection results are obtained.

The YOLOv7 series includes three basic versions: YOLOv7-tiny, YOLOv7, and YOLOv7-W6. These versions vary in the number of parameters, activation functions, network depth, and network width to meet diverse performance requirements. YOLOv7-tiny has the fewest parameters and the smallest network depth and width among all YOLOv7 versions, which makes it slightly less accurate, but it has the fastest detection speed. The detection speed

introduces CBAM [32], which better captures image similarities and differences; Section 3.4 discusses the improved self-calibrated convolutional module; and Section 3.5 introduces the EIoU loss function.

3.2. Space-to-Depth Convolution (SPD-Conv)

When the image resolution is very low or the object to be detected is small, traditional convolutional neural networks' performance will deteriorate rapidly. In reference [31], the authors point out that this is mainly because the existing CNN architecture uses pooling layers or strided convolutions, resulting in the loss of fine-grained information and less efficient feature representation. Therefore, considering that Dotted defects in cigarette appearance defects are extremely small targets and often fail to be detected by existing models, this paper introduces the SPD-Conv module into the backbone network of the YOLOv7-tiny model to replace the pooling layer (MP module) in the original network.

The SPD-Conv module consists of a space-to-depth (SPD) layer and a non-strided convolution (Conv) layer. SPD-Conv is formed by connecting SPD layers and Conv layers in series. Specifically, the input feature map is first converted from space to depth through the SPD layer, and then the output is convolved through the Conv layer [31]. This combination method can reduce the size of the feature map without losing information, retaining all the information in the channel, which greatly enhances the ability of the model to detect low-resolution images and small targets [31]. Figure 3 shows the structure of SPD-Conv.

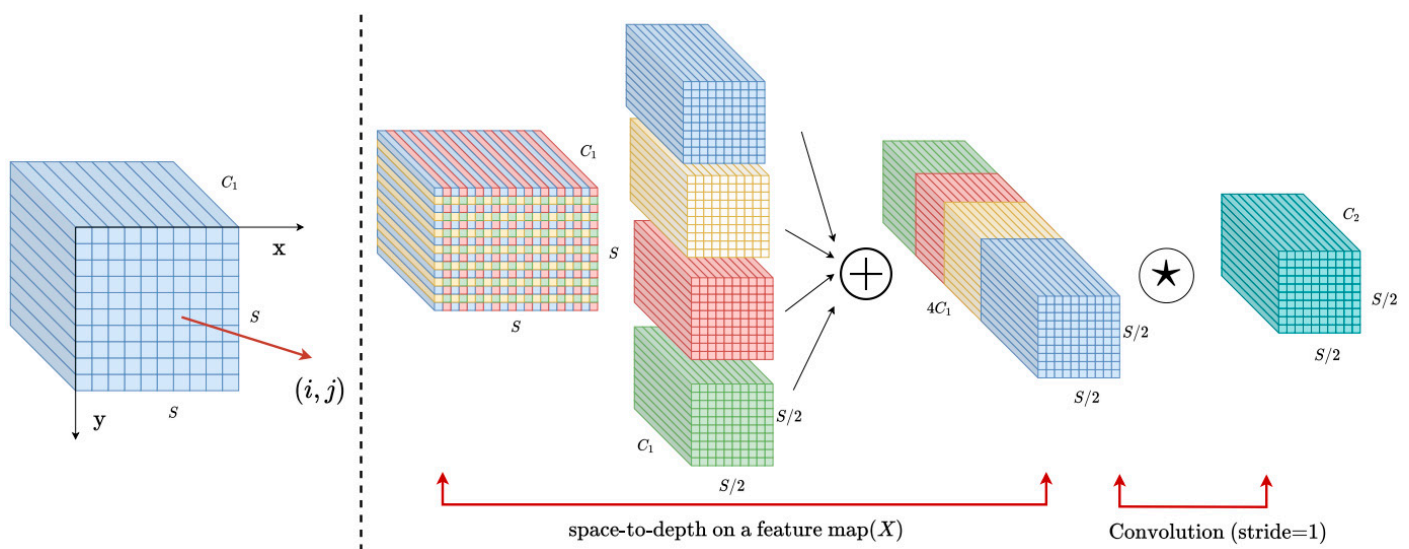


Figure 3. SPD-Conv structure.

In the backbone network of the YOLOv7-tiny model, there are three max pooling layers (MP modules) with a stride length of 2. Although these three pooling layers achieve feature dimension reduction and decrease the algorithm's computational complexity, they also cause the loss of much fine-grained information. Therefore, this paper considers replacing MP modules with SPD-Conv modules. The number of MP modules that needed to be replaced was determined by the ablation experiments, with the results are displayed in the experiments section later. They indicate that replacing the last two MP modules instead of all of them in the backbone network with SPD-Conv modules is optimal. According to the analysis, the cigarette appearance defect dataset contains very small targets (Dotted) and large targets (Unfilter). Although the SPD-Conv module is effective for small targets, it splits large targets and compromises their integrity. Thus, not all of the three MP modules in the backbone network can be replaced with SPD-Conv modules. The experimental results in the experiments section support this view.

3.3. Attention Mechanism (CBAM)

The attention mechanism can enhance the algorithm's focus on useful information while filtering out irrelevant information, thus improving detection accuracy and efficiency. Due to the extensive number of layers in the YOLOv7-tiny model's network structure, some invalid feature information is quickly introduced during feature extraction and fusion, reducing the model's ability to discriminate target features. Replacing the MP module with the SPD-Conv module in Section 3.2 quadruples the number of channels in the SPD layer. This channel increase requires the model to pay attention to more information, which diminishes its ability to focus on target information. This paper introduces CBAM behind the C5 module of the original YOLOv7-tiny model to enhance the model's perception of target features and improve detection performance.

CBAM is a lightweight attention module composed of two submodules: the CAM (Channel Attention Module) submodule and the SAM (Spatial Attention Module) submodule, which perform channel and spatial attention operations, respectively, on input features. The function of the CAM submodule is to make the model pay attention to the channel information; the model can learn the importance of each channel information through training. The function of the SAM submodule is to make the network pay more attention to the spatial location information; the network can learn the importance of each location region through training. By integrating channel and spatial attention mechanisms, CBAM helps the network focus more on target objects while ignoring irrelevant background information, thereby enhancing network interpretability. The network structure of CBAM is shown in Figure 4.

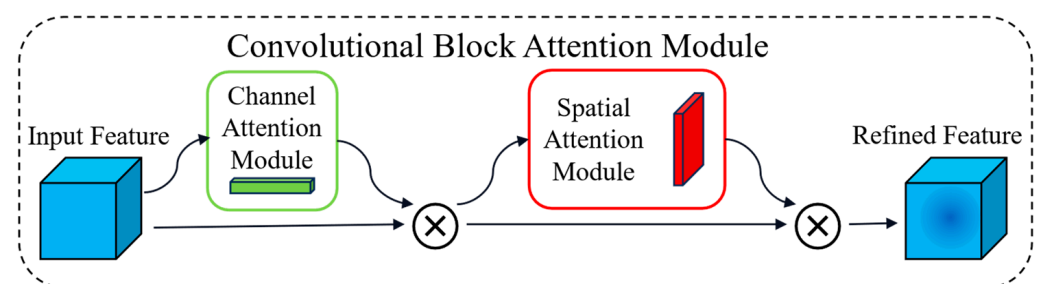


Figure 4. CBAM module structure.

3.4. Dual Self-Calibrated Convolutions (D-SCConv)

Conventional convolution has some limitations. Firstly, all of the convolution kernel learning modes of traditional convolution are similar; secondly, in the conventional convolution operation, the receptive field size of each spatial location is mainly determined by the predefined convolution kernel size, which can result in a lack of a large receptive field, thereby preventing the network from capturing sufficient high-level semantic information. Therefore, these two shortcomings may lead to low recognition of the features extracted by conventional convolution, subsequently affecting the accuracy of the model. To solve the above issues, this paper introduces self-calibrated convolutions (SCConv) [33] into the neck of the original YOLOv7-tiny model, to improve the model's receptive field for target features, thereby enhancing the detection accuracy. Additionally, SCConv can easily enhance the performance of standard convolutional layers, without introducing additional parameters and complexity, and its design is simple and plug-and-play [33]. The structure of SCConv is shown in Figure 5.

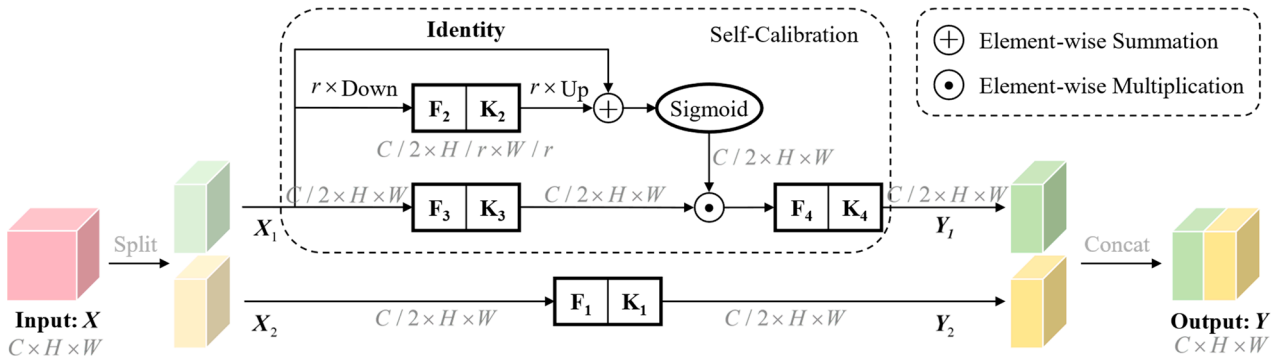


Figure 5. SCConv structure.

As shown in Figure 5, SCConv first divides the input X into two parts, X_1 and X_2 , equably and then inputs these two parts into special paths that extract different types of context information. Of the two parts, X_1 performs a self-calibration operation to obtain Y_1 . The r parameter in the self-calibration operation can adaptively adjust the receptive field of each spatial position, thereby obtaining the high-level semantic information of long-distance spatial locations and channels, enriching the diversity of output features. X_2 performs a simple convolution operation to obtain Y_2 , which preserves the contextual relations of the original space. Finally, the two intermediate outputs, Y_1 and Y_2 , are spliced together to obtain the output Y .

Some categories in the cigarette appearance defect dataset have small inter-class differences, and some appearance defects have unclear textures. This paper improves SCConv, with the expectation of further enhancing the module’s ability. The improved SCConv can adaptively construct correlations between long-distance spatial domains and channels and further expand the receptive field of each region in the network, thereby improving the feature expression ability and making the features more recognizable. The specific improvements are as follows: the part that retains the original spatial context is replaced by the self-calibration convolution block (Self-Calibration-B), and convolutions with three different convolution kernel sizes $\{K_4, K_5, K_6\}$ are used to perform the self-calibration operation on X_2 to obtain Y_2 , after which Y_1 and Y_2 are spliced to obtain the final output Y . The improved dual self-calibrated convolutions (D-SCConv) module is shown in Figure 6.

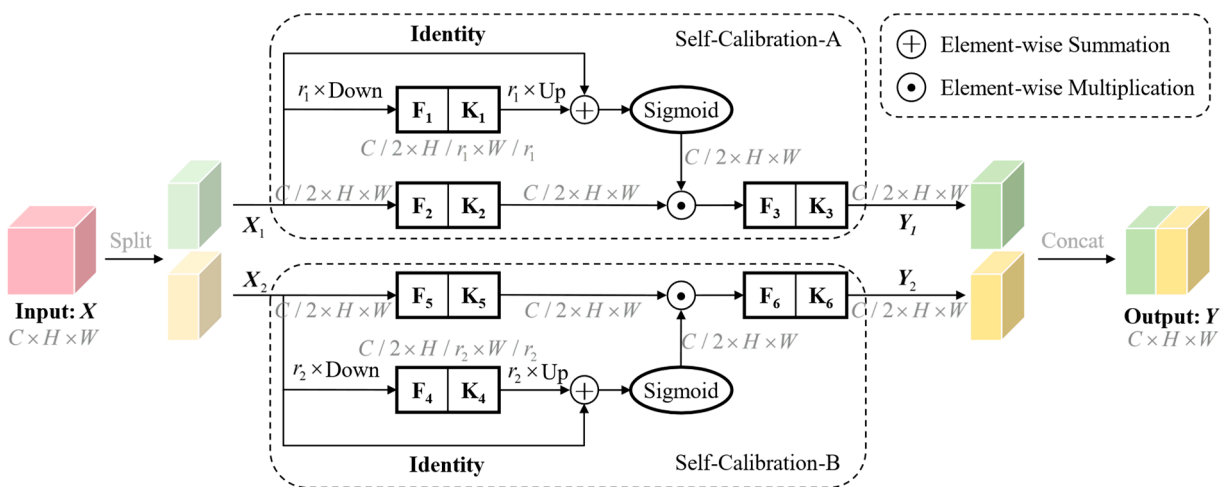


Figure 6. D-SCConv structure.

This paper conducts ablation experiments to test the improved D-SCConv module. The experimental results are displayed in the experiments section. The experiments demonstrate that using D-SCConv indeed improves the network's accuracy. Upon analysis, although SCConv is discarded to the part of retaining the original spatial context relationship, in the whole network, the output Y of D-SCConv is spliced with the output obtained by C5, CBL, and upsample. This splicing retains the spatial context relationship information, and the discarded part reduces feature redundancy. Thus, the representation learning ability of the network is further enhanced.

3.5. Improve the Loss Function

The CIoU loss function is used to calculate the loss in the YOLOv7-tiny algorithm. CIoU considers various factors—the overlap area between the predicted box and the ground truth box, the distance of the center point, the aspect ratio, etc.—and introduces a correction factor. However, CIoU has some unreasonable points, which are embodied in the following: the design of the aspect ratio is unreasonable in the sense that if the aspect ratio of the two boxes is the same, the loss of the aspect ratio is always 0; the correction factor in CIoU significantly impacts the performance of the loss function, necessitating repeated adjustments to this parameter to achieve optimal performance. To this end, EIou is introduced for loss calculation in this paper.

EIoU (Efficient IoU) is a further improvement on CIoU. It separates the aspect ratio factors of the predicted box and the ground truth box, calculating the width difference and height difference of the predicted box and the ground truth box, respectively, to replace the width and height loss based on the aspect ratio in CIoU. The calculation method for EIoU is given in Equation (1).

$$L_{EIoU} = L_{IoU} + L_{dis} + L_{asp} \quad (1)$$

By expanding Equation (1), it is possible to obtain

$$L_{EIoU} = 1 - IoU + \frac{\rho^2(b, b^{gt})}{(w^c)^2 + (h^c)^2} + \frac{\rho^2(w, w^{gt})}{(w^c)^2} + \frac{\rho^2(h, h^{gt})}{(h^c)^2} \quad (2)$$

In Equation (2), b and b^{gt} represent the predicted box and the ground truth box, respectively. $\rho^2(b, b^{gt})$ represents the Euclidean distance between the two center points of the predicted box and the ground truth box. w^c and h^c are the width and height of the minimum bounding rectangle between the predicted box and the ground truth box.

4. Experiment and Results

4.1. Dataset Introduction

The Yunnan Branch of China Tobacco Industry Company Limited provided the original images with cigarette appearance defects. In the obtained dataset, there are 1811 cigarette appearance images, including 640 with Dotted-type defects, 229 with Folds-type defects, 517 with Unfilter-type defects, 80 with Untooth-type defects, and 345 in the Normal category. Obviously, the number of samples in some defect categories is very small, indicating a problem of sample imbalance. For instance, the number of Dotted defects is eight times larger than the number of Untooth defects, making the numbers of these two types extremely unbalanced. Data augmentation techniques, for instance, flip transformation, addition of Gaussian noise, random occlusion, and random center rotation, are used to expand the dataset to make the training model more robust. After these transformations, the number of various defect images is balanced.

The augmented dataset contains 9810 cigarette appearance images. It is then divided into training, validation, and test sets in a 6:2:2 ratio, as shown in Table 1.

Table 1. Dataset division.

Type	Train	Val	Test
Dotted	1152	384	384
Folds	1096	360	376
Unfilter	1240	412	416
Untooth	1152	384	384
Normal	1242	414	414

4.2. Environment Configuration and Evaluation Index of Experimental Results

The GPU used in the experiment was an NVIDIA RTX 4090, the CPU was an Intel Core i5-13400F, the memory was 32 GB, the operating system was Windows 11, and the Python version was 3.9. The deep learning framework was PyTorch 2.0.0, and the CUDA version was 11.8. The initial learning rate was established at 0.01, with the learning rate momentum set at 0.937, the batch size at 64, and the number of epochs at 300.

The evaluation indexes used in this paper are precision (P), recall (R), mean average precision (mAP), mAP@0.5:0.95, and detection speed (FPS, frames per second). The calculation methods for these metrics are given in Equations (3)–(7).

$$P = \frac{TP}{TP + FP} \quad (3)$$

$$R = \frac{TP}{TP + FN} \quad (4)$$

In Equations (2) and (3), TP (true positive) is the number of samples of which the predicted value is positive and the true value is also positive; TN (true negative) is the number of samples of which the predicted value is negative and the true value is also negative; FP (false positive) is the number of samples of which the predicted value is positive but the true value is negative; and FN (false negative) is the number of samples of which the predicted value is negative but the true value is positive.

$$AP = \int_0^1 PRdR \quad (5)$$

$$mAP = \frac{1}{T} \sum_{i=1}^T AP(i) \quad (6)$$

In Equation (6), T is the total number of types, and $AP(i)$ is the AP for the type i .

$$mAP@0.5 : 0.95 = \frac{1}{N} \sum_{i=1}^N mAP@IoU_i \quad (7)$$

In Equation (7), N is the number of IoU thresholds, typically $N = 10$. IoU_i refers to IoU thresholds of 0.5, 0.55, 0.6, ..., 0.95 (in steps of 0.05). $mAP@IoU_i$ is the mAP for a certain IoU threshold. For instance, $mAP@0.5$ is the mAP when IoU_i is 0.5.

In summary, precision reflects the accuracy of the model prediction results; recall reflects the model's ability to identify all true positive samples; mAP is the mean AP of all types, which reflects the overall performance of the model; mAP@0.5:0.95, by considering a range of IoU thresholds, provides a more comprehensive evaluation of the model's detection performance.

4.3. Comparative Analysis of the Training Process

Figure 7 shows how the loss value changes with the number of iterations when SCS-YOLO and the original YOLOv7-tiny are trained on the same dataset under the same conditions. In Figure 7, the loss value is the total loss value, including regression, class, and confidence loss. We can see that within the first 50 epochs, the model's loss decreases

rapidly. Within the 50th to 150th epochs, the loss value gradually decreases and stabilizes. Within the 150th to 250th epochs, the loss value gradually stabilizes or even converges with the increase in the number of iterations. Therefore, 300 epochs is selected as the number of training iterations in this study.

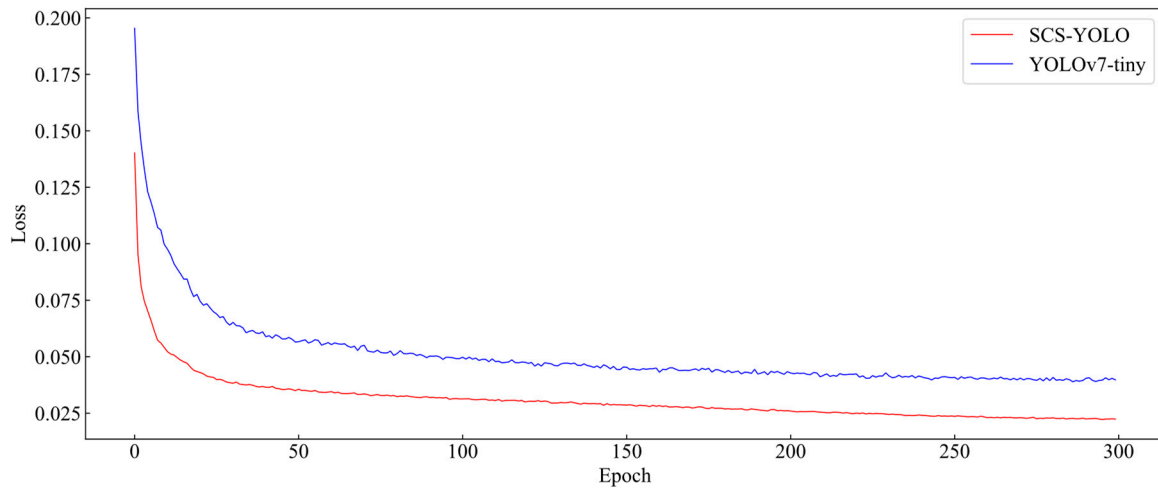


Figure 7. Comparison of loss curves.

Figure 7 shows that the loss value of SCS-YOLO decreases faster than that of the original YOLOv7-tiny and is lower for the same epoch. Therefore, SCS-YOLO performed excellently during training.

4.4. Experiments to Determine the Number of SPD-Conv

In the original YOLOv7-tiny model, there are a total of three MP modules that down-sample the feature map by a factor of two. The essence of the MP module is max pooling with a step size of two. The pooling layer can remove redundant information, reduce computational complexity, and expand receptive fields, accelerating computation speed and preventing model overfitting. However, the pooling layer inevitably loses fine-grained information during the downsampling process, leading to a decrease in the model's accuracy in small object detection. To address this problem, the SPD-Conv module is considered to replace the MP module in this paper. Corresponding comparative experiments are conducted to determine the number of substitutions, and Table 2 shows the effect when different numbers of SPD-Conv modules are added.

Table 2. Experiment on determining the number of SPD-Conv.

Experiments	Model	Precision (%)	Recall (%)	mAP@0.5 (%)	mAP@0.5:0.95 (%)
1	BASE	91.5	89.6	91.2	47.5
2	+SPD1	92.2	90.9	92.3	49.0
3	+SPD2	92.4	90.4	93.1	50.8
4	+SPD3	90.5	90.1	91.3	49.4

In Table 2, bold indicates the best results (same below), the base model is the original YOLOv7-tiny, SPD1 means that one SPD-Conv module is used to replace one MP module, and the rest are similar. It can be seen from the table that the effects of the SPD1 and SPD2 methods are greatly improved compared with those of BASE, and the SPD2 method is better. However, the effect is not significantly improved when SPD3 is used, and the accuracy is even reduced by 1% compared with that of BASE. The reason, as mentioned in Section 3.2, is that although the SPD-Conv module can significantly enhance the extraction ability of small target information, it segments large targets and reduces the global attention

ability of the model. Therefore, the SPD-Conv module cannot replace all three MP modules in the backbone network. Hence, in this paper, two SPD-Conv modules are used to replace two MP modules in the original model.

4.5. Experiments to Determine the Number of CBAM

This paper adds CBAM to the original YOLOv7-tiny model to make it focus on more critical information in the current task. Corresponding comparative experiments are conducted to determine the location and quantity of added CBAM, and the specific experimental results are shown in Table 3.

Table 3. Experiment on determining the number of CBAM.

Experiment	Model	Precision (%)	Recall (%)	mAP@0.5 (%)	mAP@0.5:0.95 (%)
1	BASE	91.5	89.6	91.2	47.5
2	+CBAM1	92.0	90.9	93.5	50.0
3	+CBAM2	92.5	91.4	93.3	50.7
4	+CBAM3	91.8	90.1	91.6	49.1
5	+SPD1+CBAM1	92.5	90.2	92.4	50.0
6	+SPD2+CBAM2	92.9	90.5	94.0	57.1
7	+SPD3+CBAM3	87.7	89.9	90.5	49.1

In Table 3, the BASE model is the original YOLOv7-tiny; CBAM1 means adding one CBAM module, and the rest are similar. From Experiments 2, 3, and 4, it can be seen that adding two CBAM modules improves the effect most significantly.

Based on this, in Table 3, Experiments 5, 6, and 7 using the SPD-Conv module and CBAM together produce better results than using the SPD-Conv module alone or the CBAM module alone. This is because after the SPD feature transformation, the features extracted by the network are segmented, and the number of channels becomes four times the original, leading to an increase in the information that the network needs to pay attention to and weakening the network's ability to concentrate on the target information. Therefore, a CBAM attention mechanism module needs to be added after the SPD-Conv module, to help the network focus on more useful information. Based on the above experiments and analysis, this paper adds two SPD-Conv modules to the improved model and a CBAM module after each SPD.

4.6. D-SCConv Comparative Experiment

This paper introduces self-calibrated convolutions into the network model to effectively expand the model's receptive field and enable the model to learn more discriminative representations. At the same time, SCConv is improved in this paper to adapt to the characteristics of the cigarette appearance image dataset. To prove whether the improved SCConv is effective, comparison experiments between the two are conducted, and Table 4 shows the results before and after the improvement.

Table 4. Comparative experiment on D-SCConv.

Experiment	Model	Precision (%)	Recall (%)	mAP@0.5 (%)	mAP@0.5:0.95 (%)
1	BASE	94.0	90.8	94.4	58.4
2	+SCConv	94.1	91.8	95.0	57.6
3	+D-SCConv	95.5	92.5	95.2	58.8

In Table 4, the BASE model is constructed by adding two SPD and two CBAM modules to the original YOLOv7-tiny, and the loss function is changed to EIoU. The results of Experiment 1 are obtained by training on the augmented dataset and subsequently testing the results. As

shown in Table 4, all performance indexes of the improved D-SCConv module have improved, with accuracy showing the most significant improvement. Thus, the D-SCConv module is introduced to the algorithm to improve the algorithm's detection capability.

4.7. Ablation Experiments

To verify the effectiveness of each improvement, a series of ablation experiments are performed using the same experimental environment and hyperparameters. Table 5 compares the various experiments on the four evaluation indexes.

Table 5. Comparison of the results of adding different modules.

CBAM	SPD	D-SCConv	EIoU	Data Augmentation	Precision (%)	Recall (%)	mAP@0.5 (%)	mAP@0.5:0.95 (%)
					91.5	89.6	91.2	47.5
✓					92.5	91.4	93.3	50.7
	✓				92.4	90.4	93.1	50.8
		✓			94.0	91.0	92.4	51.3
			✓		93.9	88.0	92.8	51.0
				✓	92.1	91.5	92.9	56.9
✓	✓				92.9	90.5	94.0	57.1
✓		✓			91.7	90.4	93.5	51.6
	✓	✓			90.1	90.0	90.8	49.8
✓	✓	✓			94.3	90.2	95.1	58.0
✓	✓	✓	✓		94.4	92.2	93.7	57.3
✓	✓	✓		✓	95.2	91.4	94.0	58.4
✓	✓	✓	✓	✓	95.5	92.5	95.2	58.8

As seen in Table 5, all improvements enhance the model's detection performance. Specifically, CBAM enables the model to focus on more useful information, increasing mAP@0.5 by 2.1%; introducing the SPD-Conv module reduces the loss of fine-grained information, improving accuracy by 0.9%; the improved D-SCConv expands the model's receptive field, significantly boosting accuracy and recall, of which the accuracy is increased by 2.5% and the recall is increased by 1.4%; the EIoU loss function accelerates model convergence and improves positioning accuracy, increasing mAP@0.5:0.95 by 3.5%; data augmentation addresses the problem of sample imbalance in the dataset, significantly improving various evaluation indexes, including a 0.6% increase in accuracy and a 9.4% increase in mAP@0.5:0.95.

4.8. Comparative Experiments

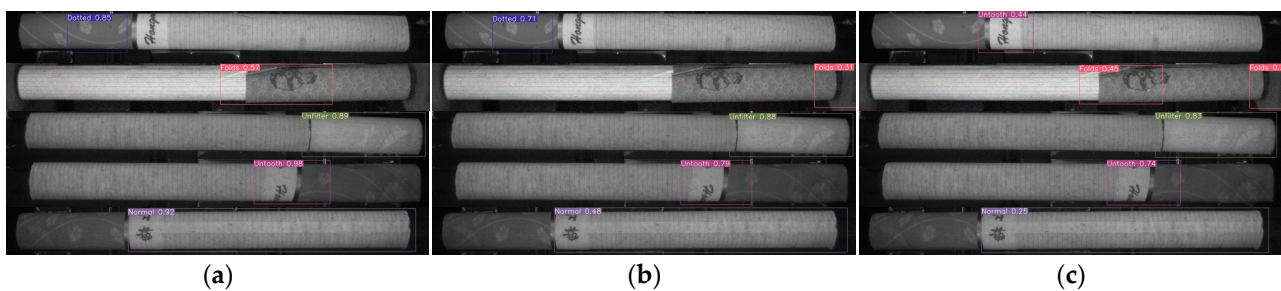
To verify the proposed algorithm's superiority, the SCS-YOLO model is compared with the Fast R-CNN, Faster R-CNN, YOLOv4, YOLOv5s, YOLOv5n, YOLOv7-tiny, YOLOv8s [34], and YOLOv9s [35] models using the main performance indexes. All detection algorithms are trained with a cigarette dataset of the same scale and the same division method, using the same equipment and parameters during training. The results of the comparative experiments are shown in Table 6.

From the results in Table 6, it is evident that the precision, recall, and mAP@0.5 of the proposed SCS-YOLO model are higher than those of the other comparison algorithms. Although its detection speed is slightly lower than that of the YOLOv7-tiny model, it is still better than other detection algorithms and can adapt to the cigarette production speed of 150–200/s on the existing production lines. Additionally, compared to the original YOLOv7-tiny model, the SCS-YOLO model shows a 4% increase in accuracy, a 2.9% increase in recall, and a 4% increase in mAP@0.5 and achieves a detection speed of 216 FPS. These results demonstrate that the proposed method excels in detection effectiveness and speed, meeting the requirements of cigarette appearance detection on the production lines.

Table 6. Comparison of the results of various algorithms.

Model	Precision (%)	Recall (%)	mAP@0.5 (%)	Speed (FPS)
Fast R-CNN	80.5	77.3	79.6	31
Faster R-CNN	82.4	81.2	83.8	36
YOLOv4	88.9	88.6	89.4	45
YOLOv5s	89.0	91.3	91.4	122
YOLOv5n	90.0	86.8	89.1	119
YOLOv7-tiny	91.5	89.6	91.2	232
YOLOv8s	92.7	89.1	92.0	187
YOLOv9s	93.1	90.6	90.5	171
Ours	95.5	92.5	95.2	216

Figure 8 shows the detection results of SCS-YOLO, YOLOv8s, and YOLOv9s on five types of defects. The proposed SCS-YOLO model accurately identifies defects in cigarette images and obtains high confidence in detecting various defects. Specifically, in the first line, SCS-YOLO correctly identifies the Dotted defect with 85% confidence, while YOLOv9s fails to detect the Dotted defect and incorrectly identifies the normal filter paper as an Untooth defect. In the second line, SCS-YOLO correctly detects the Folds defect, while both YOLOv8s and YOLOv9s misidentify shadows in the images as Folds defects. In the fifth line, due to the image being blurry and dark, the confidence of YOLOv8s and YOLOv9s is significantly lower, with YOLOv9s only 25%. However, SCS-YOLO achieves high (92%) confidence.

**Figure 8.** Visual detection comparison of different models: (a) SCS-YOLO; (b) YOLOv8s; (c) YOLOv9s.

4.9. Comparison of Cigarette Appearance Defect Detection Results

To visually illustrate the detection capability of the SCS-YOLO algorithm, Figure 9 shows the detection results of the YOLOv7-tiny algorithm and the SCS-YOLO model on five types of cigarette appearance images: Normal, Dotted, Folds, Unfilter, and Untooth.

As seen in Figure 9, the SCS-YOLO model significantly improves the degree of confidence in detecting various types of defects. Additionally, the detection box of the SCS-YOLO algorithm occupies a smaller area and is more accurate, containing as few irrelevant objects as possible, meaning that the positioning effect is better. In Figure 9c, the YOLOv7-tiny model cannot distinguish between the Folds and Untooth types, while the SCS-YOLO model learns more discriminative representations and accurately identifies the Folds type with 96% confidence. Thus, the SCS-YOLO model is more advantageous in positioning accuracy, false detection and missed detection rates, and detection confidence.

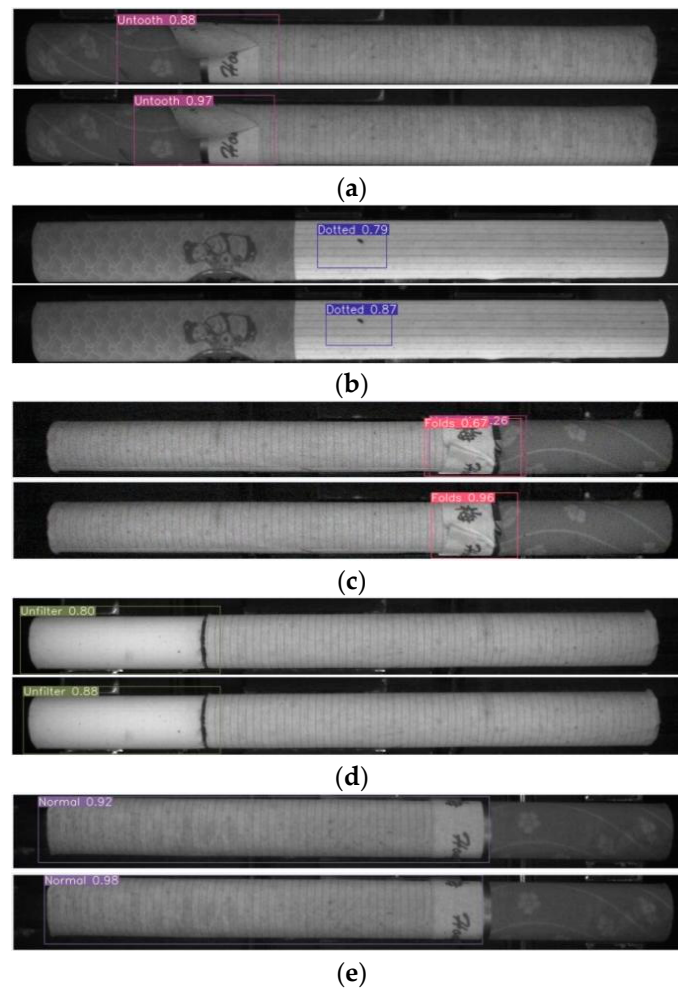


Figure 9. Visual comparison of detection results (first line: YOLOv7-tiny model, second line: SCS-YOLO model): (a) Untooth; (b) Dotted; (c) Folds; (d) Unfilter; (e) Normal.

5. Discussion and Prospects

This paper proposes a real-time and high-precision cigarette appearance defect detection model, SCS-YOLO. The primary focus is on improving the model's performance to meet the task's high accuracy and speed requirements.

Firstly, this paper uses various data augmentation techniques on the original cigarette images to address the issue of sample imbalance. Then, considering the hardware and detection speed requirements in the industrial setting, the lightweight YOLOv7-tiny network is selected as the base model. On this basis, this paper uses SPD-Conv to replace the MP module, adds attention mechanisms, expands the receptive field of the network with the improved D-SCConv, replaces the CIoU with the EIou, and names the resulting architecture the SCS-YOLO model. Experimental results demonstrate that the SCS-YOLO model is effective for cigarette appearance detection tasks, exhibiting good accuracy and recall. Furthermore, the speed satisfies the detection requirements of cigarette production lines.

Since SCS-YOLO primarily addresses three issues, namely, very small targets, significant inter-class size differences, and indistinct inter-class feature differences, it is well suited for defect detection tasks with these issues. Such tasks include aluminum profile surface defect detection, injection-molded part defect detection, steel surface defect detection, wood product defect detection, ceramic tile defect detection, etc. Additionally, SCS-YOLO is a lightweight model offering high detection accuracy and fast detection speed, making it easily adaptable to other tasks.

Although the SCS-YOLO model has significantly improved all indexes, there are still areas for further optimization. For instance, the recall is still slightly lower, with a 7.5% miss detection rate. In the future, efforts will focus on further optimizing the model. Transfer learning will then be tried to apply the model to other defect detection tasks, and integration with other quality control systems will be explored to enhance the model's applicability.

Author Contributions: Conceptualization, Y.D. and G.Y.; methodology, Y.D. and G.Y.; software, Y.D. and H.Z.; validation, Y.D., H.W. and C.M.; data curation, Y.D., H.W. and C.M.; writing—original draft preparation, Y.D. and G.Y.; writing—review and editing, Y.D., G.Y., H.Z. and H.W.; visualization, Y.D. and G.Y. All authors have read and agreed to the published version of the manuscript.

Funding: 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) and the Postgraduate Practice and Innovation Project of Yunnan University (Grant no. KC-23234471).

Data Availability Statement: We have established a dataset containing five types of cigarette appearance defects. After removing sensitive information, this dataset will be publicly released.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Liu, H.; Yuan, G. Cigarette appearance defect detection method based on improved YOLOv5s. *Comput. Technol. Dev.* **2022**, *32*, 161–167.
- Liu, H.; Yuan, G.; Yang, L.; Liu, K.; Zhou, H. An appearance defect detection method for cigarettes based on C-CenterNet. *Electronics* **2022**, *11*, 2182. [[CrossRef](#)]
- Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 580–587.
- Girshick, R. Fast r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Boston, MA, USA, 7–12 June 2015; pp. 1440–1448.
- Ren, S.; He, K.; Girshick, R.; Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. In Proceedings of the 28th Annual Conference on Neural Information Processing Systems, Montreal, QC, Canada, 7–12 December 2015.
- Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.-Y.; Berg, A.C. Ssd: Single shot multibox detector. In Proceedings of the 14th European Conference on Computer Vision—ECCV 2016, Amsterdam, The Netherlands, 11–14 October 2016; pp. 21–37.
- Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
- Redmon, J.; Farhadi, A. YOLO9000: Better, faster, stronger. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 7263–7271.
- Redmon, J.; Farhadi, A. Yolov3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.
- Bochkovskiy, A.; Wang, C.-Y.; Liao, H.-Y.M. Yolov4: Optimal speed and accuracy of object detection. *arXiv* **2020**, arXiv:2004.10934.
- Zhang, L.; Lang, X.-L.; Wang, L. Surface defect detection of aluminum profile based on image fusion and YOLOv3. *Comput. Mod.* **2020**, *8*, 8–15.
- Cheng, J.; Duan, X.; Zhu, W. Research on metal surface defect detection by improved YOLOv3. *Comput. Eng. Appl.* **2021**, *57*, 252–258.
- Wei, Z.; Xiao, S.; Jiang, G.; Wu, S.; Cheng, G. Research on surface defect detection of wood-based panels based on deep learning. *China Forest Prod. Ind* **2021**, *58*, 21–26.
- Peng, Y.; Xiao, S.H.; Ruan, J.H.; Tang, B. Research on surface defect detection of particleboard based on Faster R-CNN. *Modul. Mach. Tool Autom. Manuf. Tech.* **2020**, 91–94. [[CrossRef](#)]
- Luo, X.; Yu, F.; Peng, Y. UAV power grid inspection defect detection based on deep learning. *Power Syst. Prot. Control* **2022**, *50*, 132–139.
- Hu, X.; Yu, Z.; Liu, H.S. Weld surface defect detection method based on multi-feature extraction and BT-SVM. *Mech. Sci. Technol. Aerosp. Eng.* **2022**, *41*, 1615–1622.
- Lu, D.L.; Wang, S.Q.; Lu, H.; Zhang, P.F.; Liu, Y.F.; Yao, R.T. A defect detection method of solar cell based on improved Faster R-CNN. *Laser J.* **2022**, *43*, 50–55.
- Hu, X.; Zhou, Y.Q.; Xiao, J.; Yang, J. Surface defect detection of threaded steel based on improved YOLOv5. *J. Graph.* **2023**, *44*, 427–437.
- Wang, L.; Liu, J.; Zhang, J.; Wang, J.; Fan, X. Corn seed defect detection based on watershed algorithm and two-pathway convolutional neural networks. *Front. Plant Sci.* **2022**, *13*, 730190. [[CrossRef](#)]

20. Wan, G.; Fang, H.; Wang, D.; Yan, J.; Xie, B. Ceramic tile surface defect detection based on deep learning. *Ceram. Int.* **2022**, *48*, 11085–11093. [[CrossRef](#)]
21. Zhang, X.M.; Mao, J. Part defect detection method embedded in SENet convolutional neural network. *Agric. Equip. Veh. Eng.* **2023**, *61*, 94–98.
22. Qi, X.M.; Dong, X. Improved yolov7-tiny algorithm for steel surface defect detection. *Comput. Eng. Appl.* **2023**, *59*, 176–183.
23. Jing, J.; Wang, Z.; Rättsch, M.; Zhang, H. Mobile-Unet: An efficient convolutional neural network for fabric defect detection. *Text. Res. J.* **2022**, *92*, 30–42. [[CrossRef](#)]
24. Li, W.; Zhang, L.; Wu, C.; Cui, Z.; Niu, C. A new lightweight deep neural network for surface scratch detection. *Int. J. Adv. Manuf. Technol.* **2022**, *123*, 1999–2015. [[CrossRef](#)]
25. Qu, R.; Yuan, G.; Liu, J.; Zhou, H. 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, Xiamen, China, 22–24 October 2021; pp. 1148–1153.
26. Yuan, G.; Liu, J.; Liu, H.; Qu, R.; Zhou, H. Classification of cigarette appearance defects based on ResNeSt. *J. Yunnan Univ. Nat. Sci. Ed.* **2022**, *44*, 464–470.
27. Yuan, G.; Liu, J.; Liu, H.; Ma, Y.; Wu, H.; Zhou, H. Detection of cigarette appearance defects based on improved YOLOv4. *Electron. Res. Arch.* **2023**, *31*, 1344–1364. [[CrossRef](#)]
28. Wang, C.-Y.; Bochkovskiy, A.; Liao, H.-Y.M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 7464–7475.
29. Zhang, Y.; Ren, W.; Zhang, Z.; Jia, Z.; Wang, L.; Tan, T. Focal and efficient IOU loss for accurate bounding box regression. *arXiv* **2021**, arXiv:2101.08158. [[CrossRef](#)]
30. Zheng, Z.; Wang, P.; Ren, D.; Liu, W.; Ye, R.; Hu, Q.; Zuo, W. Enhancing geometric factors in model learning and inference for object detection and instance segmentation. *IEEE Trans. Cybern.* **2021**, *52*, 8574–8586. [[CrossRef](#)] [[PubMed](#)]
31. Sunkara, R.; Luo, T. No more strided convolutions or pooling: A new CNN building block for low-resolution images and small objects. In Proceedings of the Joint European Conference on Machine Learning and Knowledge Discovery in Databases, Grenoble, France, 19–23 September 2022; pp. 443–459.
32. Woo, S.; Park, J.; Lee, J.-Y.; Kweon, I.S. Cbam: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.
33. Liu, J.-J.; Hou, Q.; Cheng, M.-M.; Wang, C.; Feng, J. Improving convolutional networks with self-calibrated convolutions. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 10096–10105.
34. Github. Available online: <https://github.com/ultralytics/ultralytics> (accessed on 3 September 2024).
35. Wang, C.Y.; Yeh, I.H.; Liao, H.Y.M. Yolov9: Learning what you want to learn using programmable gradient information. *arXiv* **2024**, arXiv:2402.13616.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.