IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 31 March 2023, accepted 3 May 2023, date of publication 10 May 2023, date of current version 22 May 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3275023

RESEARCH ARTICLE

Few-Shot Learning for Small Impurities in Tobacco Stems With Improved YOLOv7

SHENG XUE[1], ZHENYE LI[1], RUI WU[2], TINGTING ZHU[1], YANGCHUN YUAN[2], AND CHAO NI[1]
[1]College of Mechanical and Electronic Engineering, Nanjing Forestry University, Nanjing, Jiangsu 210037, China
[2]Jiangsu Xinyuan Tobacco Sheet Company Ltd., Huaian, Jiangsu 223002, China

Corresponding authors: Tingting Zhu (tingtingzhu@njfu.edu.cn) and Chao Ni (chaoni@njfu.edu.cn)

This work was supported by the National Natural Science Foundation of China under Grant 62006120 and Grant 31570714.

ABSTRACT With the increase of public concern about health and smoking, the authorities have gradually tightened the control of tar content in cigarettes, making reconstituted tobacco a growing concern for tobacco companies. Tobacco stems are used as the main raw material for reconstituted tobacco, but they contain a large number of small broken impurities mainly from cigarette butts, which are difficult to remove efficiently by air selection and manual methods. Detection schemes for cigarette butt impurities based on computer vision and deep learning are still difficult. The scarcity of images containing foreign impurities in cigarette butts and the small size of impurities limit the efficient application of deep learning algorithms. In view of the small impurities' characteristics, this paper optimizes the model structure of the YOLOv7 algorithm, and only retains the two detection head structures with high feature resolution, which reduces the model parameters by 29.68%. Using online data augmentation and transfer learning, the difficulty of small sample datasets is overcome. After the CutMix, Mosaic, Affine transformation, Copy-paste data augmentation in this paper, the model precision is increased by 6.95%, and the recall rate is increased by 10.51%. Detection FPS has been increased from 99 FPS to 111 FPS. Precision and recall rate reached 97.21% and 92.11%. Compared with YOLOv4_csp, the precision is in-creased by 11.58%, and the recall rate is increased by 0.48%. It shows that the improved YOLOv7xs model has the potential for wide application in small target recognition. At the same time, it has shown the potential to avoid the harm of toxic substances produced by cigarette impurities in the combustion process and promotes the application of computer vision and deep learning in industrial production.

INDEX TERMS Impurities detection, deep learning, computer vision, tobacco, data augmentation, YOLO-v7.

I. INTRODUCTION

Reconstituted tobacco is a critical part of the modern tobacco industry, which is widely used to be mixed into tobacco raw materials to reduce harm to consumers' health [1]. Reconstituted tobacco is mainly composed of tobacco waste, such as soot, stems, tobacco waste and low-quality tobacco produced during the cigarette production process [2], [3]. Manufacture of reconstituted tobacco in the tobacco industry began in the 1950s, and at this stage is mainly based on the papermaking techniques, which produce reconstituted tobacco with significantly lower tar and even better quality than natural tobacco, so the technology has important social and economic value [4]. At the same time, as public health and smoking concerns increase, the relative authorities are gradually strengthening the control of tar in cigarette products, which reinforces the importance of reconstituted tobacco in cigarette production [3].

The raw materials of reconstituted tobacco are mainly scrap from cigarette production, which need to go through the storage and delivery process. Therefore, it is inevitable that the raw materials are adulterated with impurities such as metals, feathers, cigarette butts and other foreign substances. If these impurities are not clearly removed from the raw

The associate editor coordinating the review of this manuscript and approving it for publication was Vlad Diaconita.

material of reconstituted tobacco, they will produce toxic and harmful substances during the combustion process, which does not effectively reduce the amount of tobacco tar and improve the quality of cigarettes; by contrast, it will increase its harm to the human body [3].

In cigarette production and processing, the removal of impurities is a key process to improve the intrinsic quality of cigarettes, and the level of tobacco purification has an important impact on cigarette quality. There are different types of corresponding sorting methods for different types of impurities in reconstituted tobacco raw materials. First, the metal impurities are mainly removed by metal detectors. Secondly, wind selection ma-chines are used to deal with light impurities such as feathers, films, twine, weeds, and so on [1], [5]. The technology for removing impurities like tobacco's density and color, which are difficult to remove using traditional techniques, is not yet mature. For example, the use of RGB comparison of tobacco and impurities [6], or the use of hyperspectral imaging technology to detect impurities in tobacco images [7], these technologies have low adaptability and are easily affected by image noise. At present, manual sorting is still a commonly used meth-od by traditional factories to remove impurities, but it has obvious disadvantages. On the one hand, floating dust is easily generated during the sorting process of tobacco stems. Tobacco factory workers have to sort raw materials in the shop for a long time, which can easily result in long-term dust inhalation, which is dangerous to health. On the other hand, the speed of manual sorting is low and the missing rate of impurities is high [5], [8].

Due to the different stages of the tobacco processing process, impurities are introduced at different stages and need to be removed at each processing stage. Chao et al. [9]. targeted impurities generated during the opening of tobacco packages, i.e., the robotic arm opens the package of produced tobacco and tobacco shreds. Based on three cameras, color-sensitive foreign matters were identified and removed using the color classification table method (CSTM). However, the accuracy of this method is greatly affected by the location of the target and the surrounding environment due to the instability of the color value. Tang et al. [10]. proposed a cascaded foreign object detector (CFOD) based on convolutional neural networks (CNN) for the same industrial situation and achieved the localization and detection of tobacco parcels through cascaded networks. Wang et al. [8]. used YOLOX with ResNet-18 network based on deep learning for the production status of tobacco cabinets and the detection of foreign objects in the cabinets and achieved the working status of tobacco cabinets and cabinets. Recognition of foreign objects such as mops in the cabinet was realized. Thus, computer vision and deep learning approaches are gradually moving towards applications in different stages of tobacco processing.

However, the detection of impurities in the reconstituted tobacco industry field is still a challenge for computer vision and deep learning-based detection systems. This is mainly due to two factors: first, the deep learning network relies on a large amount of data to train the network, and there are few samples containing cigarette butt impurities during the processing of reconstituted tobacco, which is not enough to train a good enough net-work model; second, because cigarette butt impurities in tobacco stems, the target is small, and the RGB color and shape between positive and negative samples are very similar. For insufficient data samples, transfer learning and data augmentation are better methods to solve the problem [11], [12]. Nguyen et al. in 2020, evaluated Fast-RCNN, YOLO and other deep learning approaches in small object detection, and concluded that YOLO is a good choice without concern for training time [13]. For detection of small targets, Li et al. proposed a single-stage detection scheme to recognize small targets based on the improved YOLOv4 algorithm [14]. Zhu et al. used the improved YOLO-v5 model TPH-YOLOv5 for target detection in UAV capture scenes, overcoming difficulties such as target scale change and target object motion blur, and setting a new record for the VisDrone2021 dataset [15].

The YOLO series algorithm has become the hotspot in the field of target detection for its excellent detection performance, and YOLOv7 exceeds all known target detectors in terms of speed and accuracy in the range of 5FPS to 160FPS, with an AP of up to 56.8% [16], [17], [18], [19]. It can effectively detect impurity targets, which is important for rapid industrial development and applications.

In this paper, an improved YOLOv7 model is proposed to detect small cigarette impurities mixed in cigarette stems. According to the experiments, the proposed method not only improved the detection speed but also the classification accuracy and provides useful guidance for the subsequent process of impurities rejection [17].

The main contributions of this paper can be summarized as follows:

1) Transfer learning and optimized data augmentation methods are used to improve impurity recognition accuracy using in the case of small-scale dataset.

2) The YOLOv7 is firstly improved by optimizing model structure and parameters to detect small cigarette impurities in tobacco stems, which has achieved great performance.

II. MATERIALS AND METHODS

A. ACQUISITION OF MATERIALS

Tobacco stems are the main raw materials for reconstituted tobacco. Before entering the processing link, manual sorting is required, and lighter impurities are removed by wind separation, and metal detectors are used to peel off metal impurities. However, non-metallic impurities such as cigarette butts cannot be removed by wind selection, but only by manual sorting. However, due to the subjective nature of human decision-making, manual sorting inevitably leads to the omission of impurities. To overcome the limitations of manual sorting, it is necessary to increase the steps of cigarette impurity detection in the process of raw material impurity removal.

We installed an industrial camera above the conveyor belt that will capture tobacco raw material on the conveyor belt

from the top. The captured image data will be transmit-ted to a vision workstation where a real-time tobacco impurities detection algorithm will be performed. The data used in this paper were collected from Jiangsu Xinyuan Tobacco Flake Co., located in Jiangsu, China, which is a wholly owned subsidiary of Jiangsu China Tobacco Industry Co. In creating the dataset, a camera was installed 1.3 meters above the conveyor belt of the production line to capture the tobacco material above the conveyor belt (the light source is located below the camera). Figure 1 shows the sorting equipment and the raw material stems. The cigarette butts are mixed with the stems and are evenly distributed on the black conveyor belt. In this paper, the LUSTER line scan camera model LT-400CL-F-1 is used because it is a prism-based high-performance industrial RGB color line scan camera with three independent imagers and 4096 pixels, which can better meet the needs of small target identification, so this paper will be based on RGB images for detection and analysis.

B. TRANSFER LEARNING AND DATA AUGMENTATION

Data-hunger and data-imbalance are two major pitfalls in many deep learning approaches [20]. Because the production line still needs to complete the production task during the data set production process, and there are few sample data containing impurities during the production process, a total of 69 photos of tobacco stems and cigarette butts were created, each with 4096 × 1024 pixels. In terms of dataset expansion, this paper chooses some data augmentation methods considering that cigarette butts and cigarette stems are less different in RGB color. These methods include Cutmix [21], Mosaic [22], Affine transformation, and Copy-paste data augmentation methods [23].

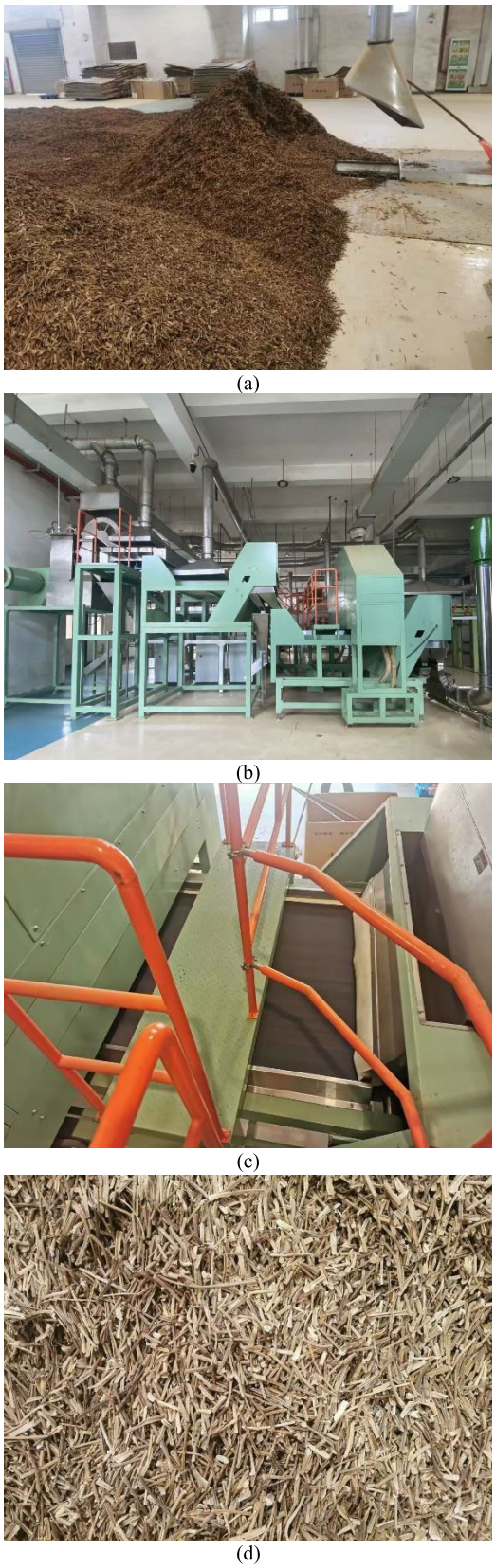
In order to overcome the difficulty of small samples, this paper not only adopts online data augmentation, but also uses transfer learning. This paper uses the pre-trained YOLOv7 model weights on the Common Objects in Context (COCO) dataset for initialization [16], [24]. This transfer learning method can effectively improve the performance of the model on small sample data, and can avoid the problem that the model is difficult to con-verge or overfit [11], [25], [26].

1) CUTMIX DATA AUGMENTATION

CutMix is a combination of CutOut and Mixup. His strategy is to cut and paste blocks between the training images and the real labels are mixed proportionally according to the area of the block [21]. The visualization is shown in Figure 2. The main purpose of using this method is to compensate for the situation that the dataset has fewer cigarette targets and less uniform distribution. Compared to the Cutout method, its ability to enrich sample data, and because the dataset discussed in this paper has only one classification, Mixup's method was not chosen for data augmentation.

2) MOSAIC DATA AUGMENTATION

The Mosaic data augmentation method is to randomly crop four images, and then stitch the four images together to obtain



(a) (b) (c) (d)

FIGURE 1. Sorting equipment and tobacco stems. (a) Tobacco stems are placed at the feed port; (b) Overall side view of the device; (c) In the image acquisition part, the smoke stem falls from the right side to the conveyor belt, and the camera and light source are sealed in the left side; (d) Tobacco stem of raw material samples for identification.

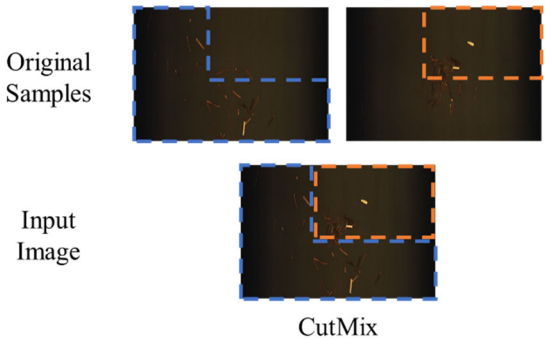


FIGURE 2. CutMix data augmentation process visualization.

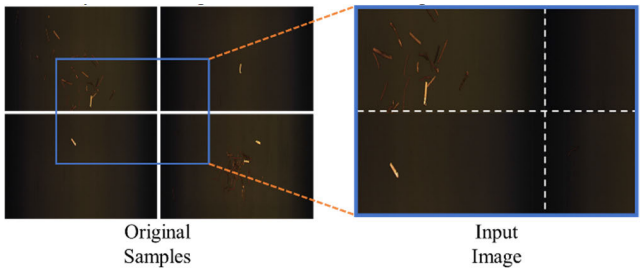


FIGURE 3. Mosaic data augmentation process visualization.

a new image, and also obtain the corresponding annotation of this image, and then we pass such a new image into the neural network to learn, which is equivalent to passing four images at once for learning. This enriches the background of the detected objects and improves the training effect by increasing the batch size in disguise.

3) AFFINE TRANSFORMATION DATA AUGMENTATION

Affine transformation refers to the process of transforming a vector space into another vector space by linear transformation, which can be realized by a compound of a series of atomic transformations, including translation, scaling, flipping, rotation and shearing transformations. Mathematically, it can be understood as a vector representation of a finite number of pixel points in an image, respectively, after affine transformation, as shown in the following equation:

$$\begin{bmatrix} X' \\ Y' \\ 1 \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & t_x \\ a_3 & a_4 & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} = A \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} \tag{1}$$

$$A = \begin{bmatrix} a & t \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & t_x \\ a_3 & a_4 & t_y \\ 0 & 0 & 1 \end{bmatrix} \tag{2}$$

where, α controls the rotation, scaling, flipping, and shearing transforms of the image. t controls the translation transform. The affine transformation is applied to replicate the changes in shape, position, etc. of the cigarette stems and butts after being squeezed.

4) COPY-PASTE DATA AUGMENTATION

Copy-paste and CutMix have similarities [23]. CutMix is to cut and paste the image blocks, while Copy-paste is to detect the image of the target instance to the new back-ground image by pasting the object that is copied and pasted, which is accurate to the pixel level, and this is the difference between it and the CutMix method. Mathematically, it can be understood as mixing two images together as followings:

$$I = I_1 \times \gamma + I_2 \times (1 - \gamma) \tag{3}$$

where, I_1 is the pasted image and I_2 is the main image, and γ is the binary mask. It is expected that more training data will be obtained by Copy-paste to replicate the random distribution of cigarette impurities.

TABLE 1. Model parameter quantity comparison.

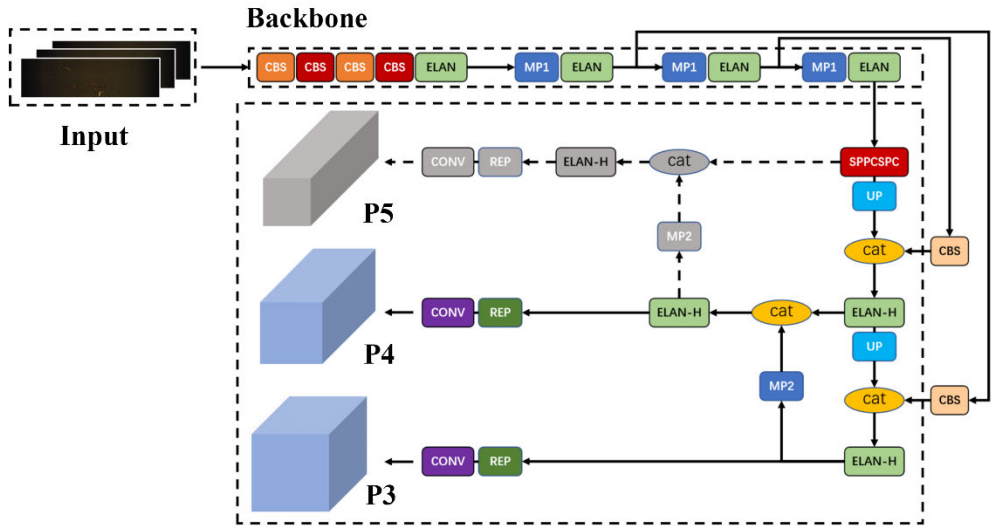
Network	Parameter quantity	GFLOPs:
YOLOv7	36481772	2.58×10^{-1}
YOLOv7xs	25972168	2.36×10^{-1}

C. IMPROVED YOLOV7 FOR DETATING CIGARETTE BUTTS

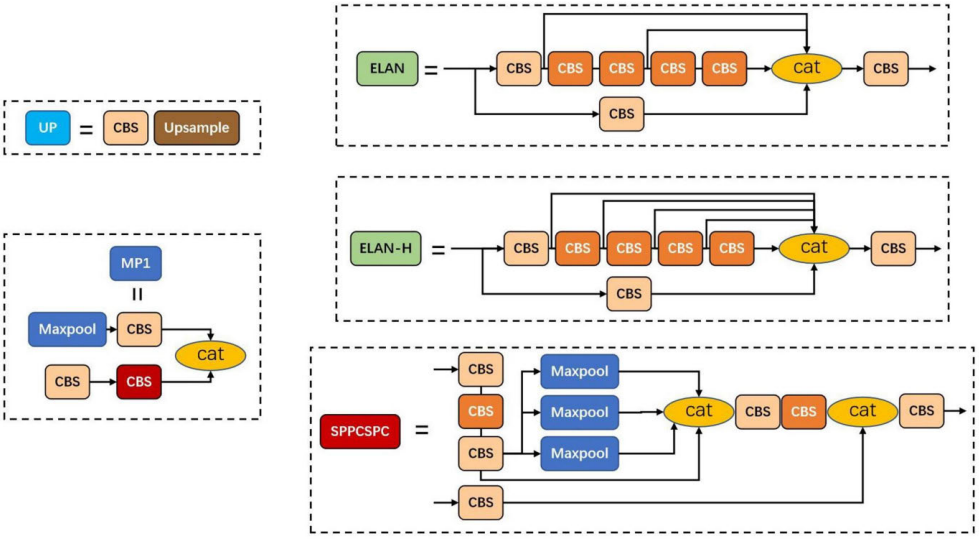
The structure of the improved YOLOv7 model is shown in Figure 4, which is mainly divided into two parts: the main structure and the basic components. The main structure consists mainly of three parts: the input, the backbone and the head. The backbone network achieves feature extraction of the input image through network units such as CBS and ELAN [16]; the head network combines the advantages of FPN and PAN and has a PA-FPN structure. The P3, P4 and P5 feature information is output from the PA-FPN structure, and the number of channels is adjusted through the Rep-Conv structure for result prediction [27].

YOLOv7 network prediction results depend on P3, P4 and P5 feature information. The resolution of the detection feature map determines the feature information contained in the feature values of a single feature layer. In order to compare the impact of different structures on the model detection structure, we improve the original model structure through feature visualization and compare it with the original model. The original YOLOv7 structure contains three detection heads, including the lower resolution P5 detection head. We believe that in the case of small targets in the dataset, down-sampling of the input image will affect the features of small targets. In the optimized network structure, we did not choose to increase the network structure of the detection heads with higher resolution, but cut off part of the network structure and kept only two detection heads. This reduces the network parameters by around 29%, and the number of models' parameters and the models' FLOPs (floating point operations) are shown in Table 1.

Impurities in cigarette stems: cigarette butts have been labelled. Figure 5 is an example of a dataset. Take this image as input and pass it into the YOLOv7 network. Using YOLOv7 and YOLOv7_xs models for object detection, respectively. In this process, the feature map is extracted from the model output and visualized for analysis. We chose the



(a) The head of the improved YOLOv7 Architecture



(b) The basic components of the improved YOLOv7 Architecture

FIGURE 4. (a) The head of the improved YOLOv7 Architecture (b) The basic components of the improved YOLOv7 Architecture The main structure of the improved YOLOv7.

FIGURE 5. Impurities in cigarette stems.

feature map corresponding to the anchor's confidence level score for visualization, because compared to other features, it can reflect the properties of the object category at the relative position.

The brighter the color, the larger the corresponding value. In Figure 6, the first three rows are features of the anchor value of the original YOLOv7 network, and the last two rows are features of the improved network. The first column of pictures represents the heat map of the coordinates of the center point of the anchor; the second column of pictures represents the width heat map of the anchor; the third column of pictures represents the height heat map of the anchor. It can be seen from the comparison that there are obvious errors in the features presented by the third row feature map P5, the first and fourth row feature maps are clearer, and the fifth row feature map most accurately represents the coordinates, width and height of the anchor where the target is located.

D. LOSS FUNCTION OF THEIMPROVED YOLOV7

Based on the prediction task above, YOLOv7 continues the loss function of YOLOv5 with object confidence loss,

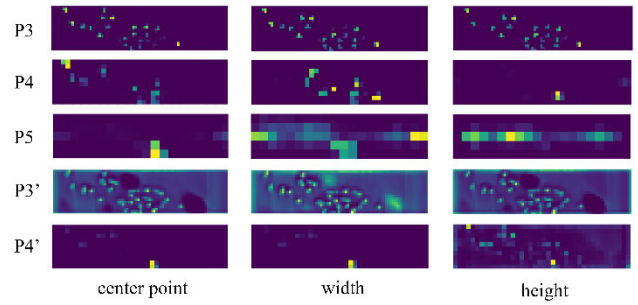


FIGURE 6. Visualize and compare the features of YOLOv7's detection head.

classification loss and coordinate loss. Coordinate loss: error between the prediction box and the calibration box, classification loss: calculate whether the anchor frame is correctly classified with the corresponding calibration, confidence loss: calculate the confidence level of the network. The relevant formula is explained as follows:

$$L_{total} = \sum(\sigma_1 \sum L_{CloU} + \sigma_2 \sum L_{obj} + \sigma_3 \sum L_{cls}) \quad (4)$$

where, N is the number of detection layers, B is the number of targets assigned to the a priori box by the label, and $S*S$ is the number of grids which the scale is separated into. L_{CloU} is the bounding box regression loss, calculated for each target; $L_{obj}$is the target object loss, calculated for each grid; L_{cls} is the classification loss, also calculated for each target σ_1, σ_2 and σ_3 are the weights of these three losses.

Classification loss and confidence loss are calculated using the BCE With Logits Loss function. The formula of the function is as follows:

$$l(x, y) = L = \{l_{1,\cdots,l_N}\}^T \quad (5)$$

$$l_n = -w_n \left[y_n \cdot \log(\sigma(x_n)) + (1 - y_n) \cdot log(1 - \sigma(x_n))\right] \quad (6)$$

$$\sigma(a) = \frac{1}{1 + exp(-a)} \quad (7)$$

where, N is the batch size, x denotes the sample, y label, a denotes the predicted output, and n denotes the total number of samples. The Complete loU loss calculation formula is shown as follows:

$$L_{CloU} = 1 - IoU + \frac{\rho^2(b, b^{gt})}{c^2} + \alpha v \quad (8)$$

$$IoU = \frac{|b \bigcap b^{gt}|}{|b \bigcup b^{gt}|} \quad (9)$$

$$v = \frac{4}{\pi^2}\left(arctan\frac{w^{gt}}{h^{gt}} - arctan\frac{w}{h}\right)^2 \quad (10)$$

where, ρ^2 is the Euclidean distance, c is the diagonal length of the smallest enclosing box covering the two boxes, and α and v are parameters. Where b^{gt} is the ground truth, and b is the predicted box.

TABLE 2. Confusion matrix for the classification results.

Labeled	Predicted	Confusion Matrix
Positive	Positive	TP
Positive	Negative	FN
Negative	Positive	FP
Negative	Negative	TN

III. EXPERIMENTAL ENVIRONMENT AND MODEL PERFORMANCE METRICS

A. EXPERIMENTAL ENVIRONMENT

The experiments in this paper are based on the deep learning framework Pytorch and run on Ubuntu 22.04. The computer's configuration is: Intel(R) Core (TM) i7-8700 CPU clocked at 3.20GHz, NVIDIA GeForce RTX 2080 Ti GPU with 12GB GPU memory and 32GB RAM.

B. EVALUATION METRICS

We used the evaluation methods of Recall Rate, Precision, mAP, and F2-score to compare different YOLOv7 model configurations. For the classification problem in this paper, the predicted and true values of the targets can be classified into the following four categories: TP (true positive), FP (false positive), FN (false negative), TN (true negative), which are summarized into the confusion matrix as shown in the following Table 2.

Recall Rate and Precision are defined as follows:

$$RecallRate = \frac{TP}{TP + FN} \quad (11)$$

$$vPrecision = \frac{TP}{TP + FP} \quad (12)$$

Recall Rate is used to measure the coverage of the model's prediction results for a given category. In this paper, this metric can be interpreted as the percentage of detected cigarette impurities in the overall cigarette impurities. Precision is used to measure the accuracy of the model's prediction results for a given category. This metric in this paper can be interpreted as the percentage of real cigarette impurities among the targets detected as cigarette impurities.

The average precision (AP) takes into account the influence of recall and precision and is the average of the highest precision under different recall conditions. Its calculation formula is as follows:

$$AP = \frac{1}{11}\sum_{r\in\{0,0.1,\ldots,1\}} P_{interp}(r) \quad (13)$$

In the evaluation of the performance of the model in this paper, we consider P and R at the same time, we should pay more attention to R, but also ensure the accuracy of P. Therefore, we introduce an F2-score metric, which considers the importance of recall to be twice as important as the importance of precision. Its formula is as follows.:

$$F2 = \frac{5 \times Precision \times Recall}{4 \times Precison + Recall} \quad (14)$$

TABLE 3. Results of ablation studies on data augmentation methods.

Index	CutMix	Mosaic	Affine-transformation	Copy-poste	Precision	Recall	F2-score
1					87.49%	79.96%	81.36%
2	✓				90.00%	89.96%	89.97%
3		✓			89.98%	90.00%	90.00%
4			✓		79.76%	80.00%	79.95%
5				✓	89.99%	90.00%	90.00%
6	✓	✓		✓	85.64%	89.98%	89.08%
7	✓	✓	✓	✓	**94.44%**	**90.47%**	**91.24%**

TABLE 4. Performance of different models.

Methods	Precision	Recall	AP	F2-score
YOLOv7	94.44%	90.47%	63.37%	91.24%
YOLOv7(d6)	94.44%	89.54%	61.95%	90.48%
YOLOv7(w6)	89.74%	**92.11%**	62.68%	91.63%
YOLOv7-tiny	93.59%	**92.11%**	60.14%	92.40%
YOLOv7xs	94.59%	92.10%	**64.83%**	92.59%
YOLOv7xs(w)	92.10%	92.06%	57.61	92.07%
YOLOv7xs(d)	**97.21%**	**92.11%**	63.09%	**93.09%**

IV. RESULTS AND DISCUSSION

A. CONTRAST EXPERIMENT RESULTS OF DIFFERENT DATA AUMENTATION METHODS

In this section, four data augmentation methods, Cutmix, Mosaic, Affine-transformation and Copy-paste, are used for ablation studies. As shown in Table 3, the method with the best performance in a single indicator is shown in bold, and the method with the best performance after using the four data augmentation methods.

In Table 3, the experimental results of Group 2 and Group 1 show that Cutmix can significantly improve the detection accuracy in the dataset of this paper; Group 3 and Group 5 compare the first group of experiments, Mosaic and Copy-paste can improve the recall rate; The comparison of Group 6 and Group 7 experiments shows that this paper combines four data augmentation methods with better performance, which can improve the detection accuracy and recall rate of the model in the case of small sample datasets.

B. PERFORMANCE OF YOLOV7 AND ITS DERIVED MODELS

In the process of model selection, different YOLOv7 series models were applied to the training and detection of the cigarette butts and tobacco stems datasets. YOLOv7, YOLOv7(d6), YOLOv7(w6), YOLOv7-tiny are the original network models of YOLOv7 series. YOLOv7xs is the model after reducing the detection head. YOLOv7xs(w) is the model after reducing the detection head and adjusting width-multiple from 1.0 to 1.3, which adjusts the network width. YOLOv7xs(d) is the model after reducing the detection head and adjusting depth-multiple from 1.0 to 1.2, which achieves the adjustment of the network depth. Finally, this paper compares the evaluation metrics such as accuracy rate, recall rate, AP and F2 score. The experimental results are shown in the following Table 4.

From Table 4, among the official YOLOv7 models, although YOLOv7-tiny is a lightweight YOLOv7 model, performs better than other models in terms of Recall and F2-score. The structure of the YOLOv7 model is modified to form three network models with different depths and widths in the next three rows. Among them, the Precision, Recall, and F2-score of YOLOv7xs(d) are the highest. Because both YOLOv7-tiny and YOLOv7xs are lightweight models, tiny has an advantage in Recall and YOLOv7xs has an advantage in AP, while the deeper YOLOv7xs network model surpasses the other models in aspects of Precision and F2-score, and is second only to the YOLOv7 model effect in AP.

C. PERFORMANCE OF DIFFERENT MODELS

Comparing the other six related object detection methods with our proposed YOLOv7xs(d), which are Efficient-det, Faster-RCNN and YOLO series models, it can be seen that YOLOv7xs(d) is superior to other algorithms in precision, recall, AP and F2-score, and FPS is second only to YOLOv3. After optimization, the YOLOv7xs(d) model achieves a good balance between F2-score and FPS, which improves the detection speed of the model while satisfying the detection accuracy.

Finally, the YOLOv7xs(d) model structure is selected for training the dataset in this paper. Compared with the YOLOv7 model, the model reduces one detection head and sets the model depth multiple to 1.2. The detection precision reaches 97.21%, which is 2.77% higher than that of YOLOv7. Recall reaches 92.11%, which is 1.64% higher than that of YOLOv7. Figure 7 shows the detection results of YOLOv7 and YOLOv7xs for impurities in cigarette butts.

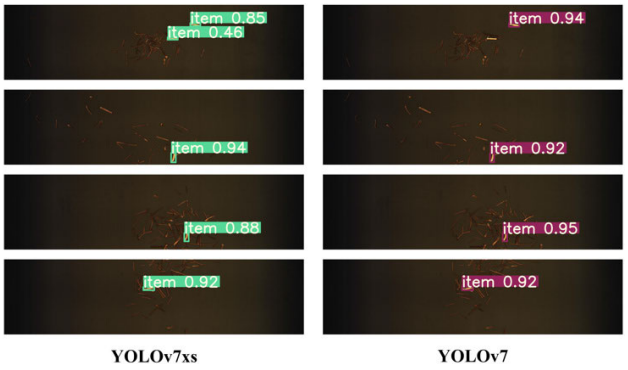


FIGURE 7. Detection Results of Cigarette Butt Impurities in Black Conveyor Belt.

V. CONCLUSION

In this study, this paper builds a small-scale dataset based on high-resolution RGB images. For the study of impurity detection in reconstituted tobacco, the task perfectly meets the requirements of practical applications and provides data support. For the ap-plication of real industrial scenarios, the YOLOv7 object detection model is selected and optimized in this study to enhance the model, and the best model is selected based on F2-score. Meanwhile, in this paper, based on a small-scale dataset, the data augmentation methods of CutMix, Mosaic, Affine transformation and Copy-paste can achieve good detection results in few-shot object detection tasks.

Impurities detection based on deep learning and computer vision has been widely used. However, there are not many applications in the field of raw materials for reconstituted tobacco. In particular, the density and color of cigarette butts are very close to tobacco stems. It is very difficult, whether it is manual sorting or machine rejection. We choose the popular YOLOv7 model, and optimize the model parameters and model structure to overcome the difficulties of small impurities and small sample datasets, which have a very great impact on achieving good results in the experiment. This project pursues recognition speed as well as recognition accuracy and recall rate, which provides a good basis for impurity rejection in practical applications.

Reconstituted tobacco with significantly lower tar and even better quality than natural tobacco. In this paper, the YOLOv7 model is introduced for the first time to detect im-purities in tobacco based on RGB images, mainly for the removal of impurities from raw materials of reconstituted tobacco. It also has a positive impact on the detection of other impurities, promoting the rejection of impurities in the raw material production process and avoiding the secondary harm of toxic substances produced during the combustion of impurities.

Due to the limitation of the actual production situation, our dataset suffers from a small sample size and unbalanced positive and negative samples. Although we compensated with data augmentation to achieve better experimental results. However, in the future, we will continue to expand and improve our dataset. In the meantime, we will focus on more application scenarios for small impurities and few-shot objects detection.

ACKNOWLEDGMENT

(Sheng Xue and Zhenye Li contributed equally to this work.)

REFERENCES

[1] W. Wang, Y. Wang, L. Yang, B. Liu, M. Lan, and W. Sun, "Studies on thermal behavior of reconstituted tobacco sheet," *Thermochim. Acta*, vol. 437, nos. 1–2, pp. 7–11, Oct. 2005, doi: 10.1016/j.tca.2005.06.002.

[2] M. Chen, Z. Xu, G. Chen, H. Wang, C. Yin, Z. Zhou, W. Sun, Y. Li, and F. Zhong, "The influence of exogenous fiber on the generation of carbonyl compounds in reconstituted tobacco sheet," *J. Anal. Appl. Pyrolysis*, vol. 105, pp. 227–233, Jan. 2014, doi: 10.1016/j.jaap.2013.11.008.

[3] L. Wang, Y. B. Wen, D. P. Sun, Y. Mao, and Y. J. Yao, "Study on the decrease of harmful substance in paper-process reconstituted tobacco sheet," *Adv. Mater. Res.*, vols. 314–316, pp. 2338–2343, Aug. 2011, doi: 10.4028/www.scientific.net/AMR.314-316.2338.

[4] T. H. Huang, Q. S. Shi, K. Wei, J. P. Gui, and L. S. Zheng, "Effect of enzyme on harmful components reduction in reconstituted tobacco," in *Proc. Int. Workshop Mater., Chem. Eng. (IWMCE)*. SciTePress-Science and Technology Publications, 2018, pp. 351–356.

[5] Z. K. Li, Y. J. Fan, Y. S. Zou, M. Y. Wu, and G. Y. Liu, "Study and application of impurity removal methods in tobacco production," *Adv. Mater. Res.*, vols. 1049–1050, pp. 1131–1134, Oct. 2014, doi: 10.4028/www.scientific.net/AMR.1049-1050.1131.

[6] C. Kai, X. Qian, X. Bo, M. Chao, and Z. Z. Wei, "A machine vision algorithm for foreign bodies detection in tobacco conveyor," in *Proc. Int. Conf. Sens. Instrum. IoT Era (ISSI)*, Lisbon, Portugal, Aug. 2019, pp. 1–6, doi: 10.1109/ISSI47111.2019.9043657.

[7] Z. Long, M. Xiaoyu, L. Zhigang, and L. Yong, "Application of hyperspectral imaging technology in classification of tobacco leaves and impurities," in *Proc. 2nd Int. Conf. Saf. Produce Inf. (IICSPI)*, Chongqing, China, Nov. 2019, pp. 157–160, doi: 10.1109/IICSPI48186.2019.9095975.

[8] C. Wang, J. Zhao, Z. Yu, S. Xie, X. Ji, and Z. Wan, "Real-time foreign object and production status detection of tobacco cabinets based on deep learning," *Appl. Sci.*, vol. 12, no. 20, p. 10347, Oct. 2022, doi: 10.3390/app122010347.

[9] M. Chao, C. Kai, and Z. Zhiwei, "Research on tobacco foreign body detection device based on machine vision," *Trans. Inst. Meas. Control*, vol. 42, no. 15, pp. 2857–2871, Nov. 2020, doi: 10.1177/0142331220929816.

[10] J. Tang, H. Zhou, T. Wang, Z. Jin, Y. Wang, and X. Wang, "Cascaded foreign object detection in manufacturing processes using convolutional neural networks and synthetic data generation methodology," *J. Intell. Manuf.*, pp. 1–17, Jun. 2022, doi: 10.1007/s10845-022-01976-3.

[11] R. Barman, S. Deshpande, S. Agarwal, U. Inamdar, M. Devare, and A. Patil, "Transfer learning for small dataset," in *Proc. Nat. Conf. Mach. Learn.*, Berlin, Germany, 2019, pp. 1–6.

[12] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *J. Big Data*, vol. 6, no. 1, pp. 1–48, Dec. 2019.

[13] N.-D. Nguyen, T. Do, T. D. Ngo, and D.-D. Le, "An evaluation of deep learning methods for small object detection," *J. Electr. Comput. Eng.*, vol. 2020, pp. 1–18, Apr. 2020, doi: 10.1155/2020/3189691.

[14] F. Li, D. Gao, Y. Yang, and J. Zhu, "Small target deep convolution recognition algorithm based on improved YOLOv4," *Int. J. Mach. Learn. Cybern.*, vol. 14, no. 2, pp. 387–394, Feb. 2023, doi: 10.1007/s13042-021-01496-1.

[15] X. Zhu, S. Lyu, X. Wang, and Q. Zhao, "TPH-YOLOv5: Improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops (ICCVW)*, Montreal, QC, Canada, Oct. 2021, pp. 2778–2788, doi: 10.1109/ICCVW54120.2021.00312.

[16] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," Jul. 2022, *arXiv:2207.02696*. Accessed: Dec. 1, 2022.

[17] K. Jiang, T. Xie, R. Yan, X. Wen, D. Li, H. Jiang, N. Jiang, L. Feng, X. Duan, and J. Wang, "An attention mechanism-improved YOLOv7 object detection algorithm for hemp duck count estimation," *Agriculture*, vol. 12, no. 10, p. 1659, Oct. 2022, doi: 10.3390/agriculture12101659.

[18] J. Zhou, Y. Zhang, and J. Wang, "A dragon fruit picking detection method based on YOLOv7 and PSP-ellipse," *Sensors*, vol. 23, no. 8, p. 3803, Apr. 2023, doi: 10.3390/s23083803.
[19] Y. Zhang, Y. Sun, Z. Wang, and Y. Jiang, "YOLOv7-RAR for urban vehicle detection," *Sensors*, vol. 23, no. 4, p. 1801, Feb. 2023, doi: 10.3390/s23041801.
[20] R. Wang, S. Hoppe, E. Monari, and M. F. Huber, "Defect transfer GAN: Diverse defect synthesis for data augmentation," 2023, *arXiv:2302.08366*.
[21] S. Yun, D. Han, S. Chun, S. J. Oh, Y. Yoo, and J. Choe, "CutMix: Regularization strategy to train strong classifiers with localizable features," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 6023–6032.
[22] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal speed and accuracy of object detection," 2020, *arXiv:2004.10934*.
[23] G. Ghiasi, Y. Cui, A. Srinivas, R. Qian, T.-Y. Lin, E. D. Cubuk, Q. V. Le, and B. Zoph, "Simple copy-paste is a strong data augmentation method for instance segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 2918–2928.
[24] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft COCO: Common objects in context," in *Computer Vision—ECCV 2014*. Zurich, Switzerland: Springer, Sep. 2014, pp. 740–755.
[25] C. Wang and Z. Xiao, "Potato surface defect detection based on deep transfer learning," *Agriculture*, vol. 11, no. 9, p. 863, Sep. 2021, doi: 10.3390/agriculture11090863.
[26] L. Deng, J. Li, and Z. Han, "Online defect detection and automatic grading of carrots using computer vision combined with deep learning methods," *LWT*, vol. 149, Sep. 2021, Art. no. 111832, doi: 10.1016/j.lwt.2021.111832.
[27] X. Ding, X. Zhang, N. Ma, J. Han, G. Ding, and J. Sun, "RepVGG: Making VGG-style ConvNets great again," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 13733–13742.

SHENG XUE was born in Huaian, Jiangsu, China, in 2000. He received the bachelor's degree in engineering from Nanjing Forestry University, in 2022, where he is currently pursuing the master's degree. His research interests include the application of intelligent algorithms in industry and image data processing.

ZHENYE LI was born in Yangzhou, Jiangsu, China, in 1997. He received the bachelor's degree in engineering from Nanjing Forestry University, in 2019, where he is currently pursuing the master's degree. His research interests include the application of intelligent algorithms in industry, image data processing, and hyperspectral analysis.

RUI WU was born in Huaian, Jiangsu, China, in 1976. He received the B.S. degree from Changzhou University formerly known as the Jiangsu Institute of Petrochemical Technology, in 1998, and the master's degree in mechanical automation from the University of Science and Technology of China, in 2005. He is currently an Engineer and Jiangsu Xinyuan Tobacco Sheet Company Ltd., Huaian. His research interest includes artificial intelligence in industrial application, especially in the field of tobacco.

TINGTING ZHU received the Ph.D. degree in pattern recognition and artificial intelligence from the School of Automation, Southeast University, in 2019. She is currently with the College of Mechanical and Electronic Engineering, Nanjing Forestry University, China. She achieved a fellowship jointly awarded by Fonds de Recherche du Quebec—Nature et Technologies (FRQNT) and the China Scholarship Council and studied as a Visiting Student with the Department of Atmospheric and Oceanic Sciences, McGill University, Canada, from 2017 to 2018. Her current research interests include machine learning, data processing and modeling, renewable energy generation forecast, and climate feedback.

YANGCHUN YUAN was born in Rucheng, Hunan, China, in 1984. He received the B.S. degree in chemistry and industry of forest products from Southwest Forestry University, in 2008, and the master's degree in pulp and paper engineering from Nanjing Forestry University, in 2011. Currently, he is the Deputy Director of the Production Management Department, Jiangsu Xinyuan Tobacco Sheet Company Ltd. His research interests include production technology and application of tobacco reconstituted raw materials.

CHAO NI was born in Nanjing, Jiangsu, China, in 1979. He received the B.S. degree in automation from the Nanjing University of Science and Technology, Nanjing, in 2001, and the Ph.D. degree in control theory and control engineering from Southeast University, Nanjing, in 2008. He is currently an Associate Professor with the Automation Department, Nanjing Forestry University, China. From October 2017 to November 2018, he was a Visiting Scholar with the University of Maryland, College Park, USA. His research interests include artificial intelligence in industrial application, data processing, and spectroscopy analysis.