A Deep Learning-based Method for Production Carton Packaging Defect Detection

Zhimin Lu, Xueliang Yang, Zhibin Zhang, Jingqiang Jiang*, Wencan Li, Weiwei Zhang, Mengling Huang
Longyan Tobacco Industry Co., Ltd.,
No. 1299 Shengfeng Rd, Xinluo District, Longyan 364021, China
*1012240592@qq.com; yxl23077@fjtic.cn; zzb23155@fjtic.cn; *jjq22102@fjtic.cn; lwc22355@fjtic.cn; zww22404@fjtic.cn; hml23122@fjtic.cn*

***Abstract*—In the automation production packing system, appearance defects on the carton packaging films occur frequently. To solve this problem, a deep learning-based method is proposed in this work to identify film defects during the packaging process to ensure the production packaging quality. This method adopts the deformable convolution network (DCN) v2-C3 module to replace the C3 module in the Neck part of the You Only Look Once version 5 (YOLOv5s) network to extract the deep feature information of the defects in the production carton packaging, with the purpose to improve the spatial transformation ability of the detection model and the model generalization ability to different shapes of targets. Field data are used to evaluate the proposed method. The analysis results indicate that the recognition rate of the proposed method is 99.3 % for different carton packaging defects; and compared to the original YOLOv5s method, the detection accuracy of the proposed method increases by 2.7 % and the scrap rate is reduced by 1.3 %. As a result, the proposed method can meet the requirements for defect detection in the production carton packaging in practical applications.**

***Index Terms*—Intelligent manufacturing; Carton packaging; Defect detection; Deep learning.**

I. INTRODUCTION

Under the development of intelligent manufacturing transformation in production processing, the integration and intelligence level of production packaging lines continues to improve. However, detecting surface defects in the carton packaging in continuous high-speed production remains a common technical challenge in industry [1]. Common packaging defects observed during on-site production include a missing transparent film, a misaligned transparent film pull line, a torn transparent film, and folded corners of the transparent film. These visual flaws directly reduce product yield, increase material waste and rework costs, and constrain overall production efficiency. Therefore, online inspection of carton packaging integrity is an essential core process within the quality control system of smart manufacturing.

In actual production processes, adapting to automation, intelligence, and digitalization of entire production lines makes defect detection in production outer packaging an essential step. From an economic perspective, a 2 % improvement in detection accuracy could reduce the scrap rate by approximately 1 %, saving enterprises substantial costs, minimizing raw material waste, and generating significant economic benefits. Furthermore, improving detection efficiency and accuracy can further optimize production processes and improve overall productivity [2].

Under the high-speed operation conditions of intelligent manufacturing for carton packaging lines, typical types of defects, missing transparent film, misaligned pull tab, film damage, and folded corners, exhibit significant size variations and strong reflective interference from the film. Traditional image processing algorithms rely on manually designed features, resulting in poor generalization and high rates of false positives and missed detections. The You Only Look Once version 5 (YOLOv5s) lightweight single-stage object detection model offers both fast inference speed and multi-class classification capability, meeting the real-time online inspection requirements of production lines. However, the original model struggles to detect small or low-contrast transparent film defects. Therefore, building a detection model for multiple surface defects in carton packaging by enhancing feature extraction and optimizing the loss function based on the YOLOv5s represents a feasible technical approach to address common industry inspection challenges and improving the quality control system in smart factories. Nevertheless, in carton packaging detection, related work is very limited, although these improved YOLOv5s models have been reported in the existing literature in other applications (not exactly in carton packaging detection).

To address the packaging inspection issues mentioned above, a deep learning-based method is proposed by incorporating a deformable convolution network (DCN) v2-C3 module into the YOLOv5s network to optimize the extraction of deep feature information for packaging film defects, thus improving the accuracy of defect detection and reducing production waste. The results of the experimental evaluation demonstrate a good detection performance of the proposed method.

This work is organised as follows. The related works are introduced in Section II, and Section III presents the proposed YOLOv5s method. Section IV evaluates the performance of the proposed method using field data and analyses the test results. Section V concludes the main findings of this study.

II. RELATED WORKS

In recent years, with the rapid development of deep

Early Access

learning, particularly centered on deep convolutional neural networks, new solutions have emerged to identify defects in production carton packaging. For example, Zhao *et al.* [3] developed a packaging appearance inspection system that used image texture analysis and template matching techniques to detect packaging surface defects. Yan [4] constructed an anomaly detection model based on visual perception feature analysis with a multi-classifier fusion strategy to improve the packaging detection robustness. Cai *et al.* [5] developed a visual inspection system for packaging machines, where Haar wavelet feature extraction is combined with pattern matching technology to achieve automated defect detection. Song, Wang, and Lou [6] developed a Single Shot MultiBox Detector (SSD)-based detection method to detect carton packaging defects. A self-built data set was used to evaluate the SSD-based method to reach an average accuracy of 96.62 %. Wu and Lu [7] used support vector machines (SVM) in the visual system for the packing box defect to achieve a detection rate of 98.0296 %. Sun [8] proposed a method to detect production packaging defects using the SVM, where template matching was used to locate the inspection region and the Haar wavelet transform was used to extract texture features from the time-frequency domain images. Meanwhile, Yang, Han, Tang, Li, and Luo [9] combined SVM with edge computing for logistics package defects. Chen *et al.* [10] designed a production packaging defect detection system based on the YOLOv4 convolutional neural network (CNN). After applying filtering to the production packaging images, the YOLOv4 architecture-based detection model leverages a multi-layer feature fusion mechanism to improve the defect recognition accuracy. Li *et al.* [11] also used YOLOv4 in an online detection system for production packaging defects and produced an average detection accuracy of 95.447 % on different packaging defects. Liang *et al.* [12] adopted YOLOv8 in a three-dimensional (3D) laser vision system for carton packaging monitoring. Pan, Yang, Liu, and Lv [13] presented an oriented R-CNN method for product packaging detection and found its superiority over several CNN versions.

As can be seen in the related works on the detection of production packaging defects, currently artificial intelligence (AI) techniques have been introduced in the carton packaging defect detection [14]. However, to meet online detection during the packaging process, the defect detection system requires quick response and precise identification performance. Existing AI techniques, including SSD, SVM, CNN, and YOLO, have different advantages and disadvantages in the packaging defect detection. Table I compares these popular AI methods in packaging defect detection.

As can be seen in Table I, the four different methods present different advantages and disadvantages, which allows them to be applied to different application scenarios. Accordingly, the SSD is suitable for industrial applications that need to balance the detecting speed and accuracy; the SVM is a good choice for small-sample and high-dimensional classification tasks; the CNN is effective for basic tasks such as image feature extraction and classification/segmentation; the YOLO is a proper selection for object detection scenarios with extremely high real-time requirements.

In the detection of production packaging defects, a basic requirement of online detection is the computation speed, and the second is the detection accuracy. With this in mind, the YOLO is among the best suitable options in this study. In the existing literature, YOLOv4 [10], [11] and YOLOv8 [12] have been introduced in the packaging defect detection; however, in the YOLO series, YOLOv5s has very distinct characteristics in terms of performance, deployment difficulty, and applicable scenarios due to its lightweight and fast computing speed. As a result, to meet the online detection requirement, this study adopts the YOLOv5s as the core detector for production packaging defects.

TABLE I. AI IN PACKAGING DEFECT DETECTION.

Methods	Advantages	Disadvantages
SSD	Classification in single forward propagation; suitable for large targets with excellent accuracy; well-developed industrial ecosystem	Sensitive to background noise; poor accuracy in small target detection; high storage overhead; requiring additional hard example mining strategies
SVM	Strong generalization ability; stable performance on small sample and high-dimensional datasets; good robustness to noise	High computational complexity; high parameter tuning costs; complex logic for native multi-class classification
CNN	Significantly reduces model parameters; high degree of parallelism; automatically extracts features; excellent generalization and transfer capabilities	Sensitive to input image size; requires a large amount of labeled training data; suffers from edge information loss
YOLO	Extremely fast inference speed; low false detection rate; low deployment threshold and a well-developed industrial ecosystem	Early versions had limited accuracy in detecting small objects; generalization ability on small datasets is not as good as SVM

The main contributions of this work include the following.

1. An online detection system is developed based on YOLOv5s for production packaging defects to increase the efficiency and quality control of production lines.
2. An improved YOLOv5s detector is proposed by incorporating a DCN v2-C3 module to optimize the feature extraction for the production packaging defects, thus improving the detection accuracy.
3. The present online detection system is evaluated using a real-world production line, and the online detection rate is 95.15 % for the production packaging defects.

III. PROPOSED DEEP LEARNING METHOD

A. YOLOv5s Object Detection

In the evolution of object detection technology, the fifth-generation YOLOv5 architecture is based on the third-generation YOLOv3 framework and incorporates the optimization strategies of fourth-generation YOLOv4 [15]. The YOLOv5 architecture excels in real-time processing, detection accuracy, and engineering deployment convenience, achieving significant improvements over previous generations in key performance metrics. The YOLOv5 architecture provides five parameter-scale implementations, including YOLOv5n, YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x. Among them, the YOLOv5s is the lightweight model in the YOLOv5 series and

Early Access

is able to balance detection accuracy and inference speed. Hence, this study selects YOLOv5s (v7.0) as the baseline architecture for carton packaging detection.

As shown in Fig. 1, the YOLOv5 network model consists of three core components, including the backbone, neck, and head [16]. The backbone encodes the target features through multi-level abstraction using the CBS (convolution-batch normalization-activation) units, C3 (CSP Bottleneck with three convolutions) modules, and spatial pyramid pooling-fast (SPPF) modules. The neck adopts an improved path aggregation network (PANet) structure, enabling cross-scale interaction between deep and shallow features via a bidirectional feature pyramid, effectively mitigating feature loss and enhancing the model ability to extract multi-scale features. The detection head serves as both a classifier and a regressor for the YOLOv5 model using three parallel branches that perform the coordinate regression and class prediction on the feature maps at three scales (i.e., 8×8, 16×16, and 32×32).

As can be seen, in the Backbone the C3 module mainly serves to deepen the network and reduce the number of parameters. The C3 module can divide the input feature map into two parts: one part first uses a 1×1 convolution for dimensionality reduction, followed by a 3×3 convolution for feature extraction. Compared to directly using a 3×3 convolution for feature extraction, the C3 approach effectively reduces computational load. Finally, the two parts are concatenated and passed through another convolution to produce the output of the C3 module.

The SPPF module replaces three parallel max-pooling operations with a serial arrangement, and each max-pooling operation uses a 5×5 pooling kernel. This approach significantly reduces the computational load while maintaining the detection accuracy.

In the PANet module, first, the feature map of the last layer of the input image with a resolution of 20×20×512 is compressed into channels through a 1×1 convolution and upsampled, then concatenated with the 40×40×256 resolution feature map. After that, through feature fusion and sampling in the C3 module, feature information can be preserved to a great extent. Next, the 40×40×256 resolution feature map undergoes convolution and upsampling, then is stacked with the 80×80×128 resolution feature map. After another C3 module operation, the first stacked feature layer output is obtained. Finally, the first stacked feature layer output is downsampled and stacked with the feature from the previous layer on the same scale, repeatedly using the C3 module operation to obtain the second stacked feature layer. The second stacked feature layer then undergoes the same steps sequentially to obtain the third output layer. As a result, the PANet module can concatenate shallow feature information into deeper feature maps, enriching the extracted feature information and preventing the loss of shallow features caused by the increased network depth.

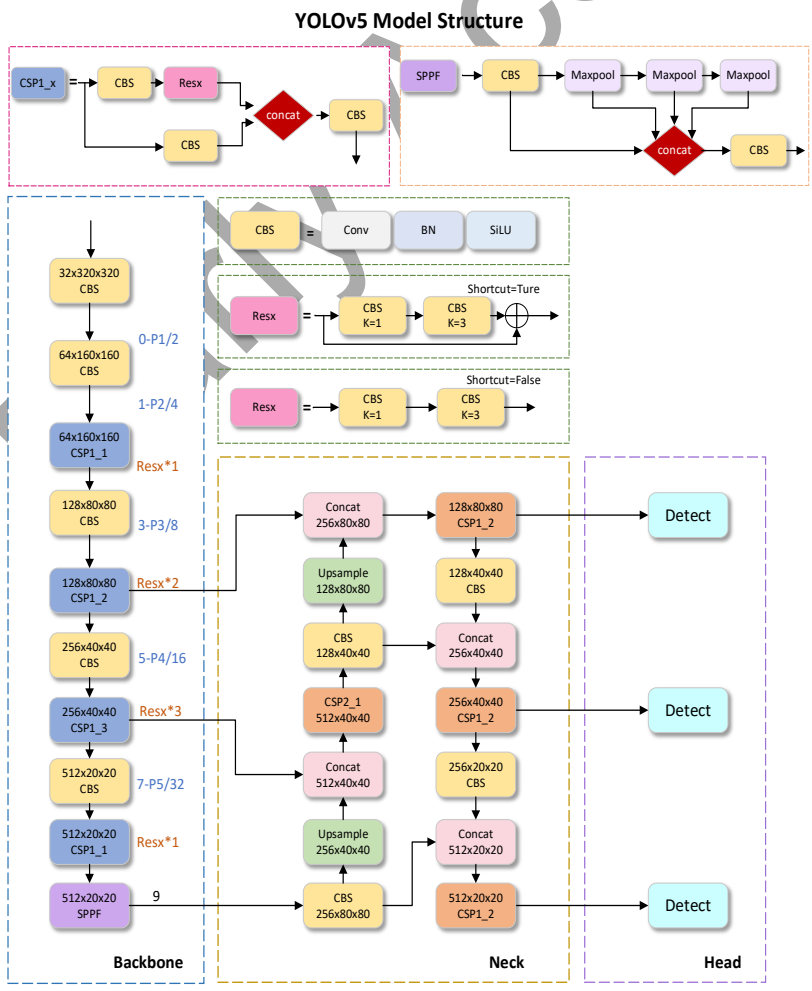


Fig. 1. Developed YOLOv5s network model for packaging detection [15].

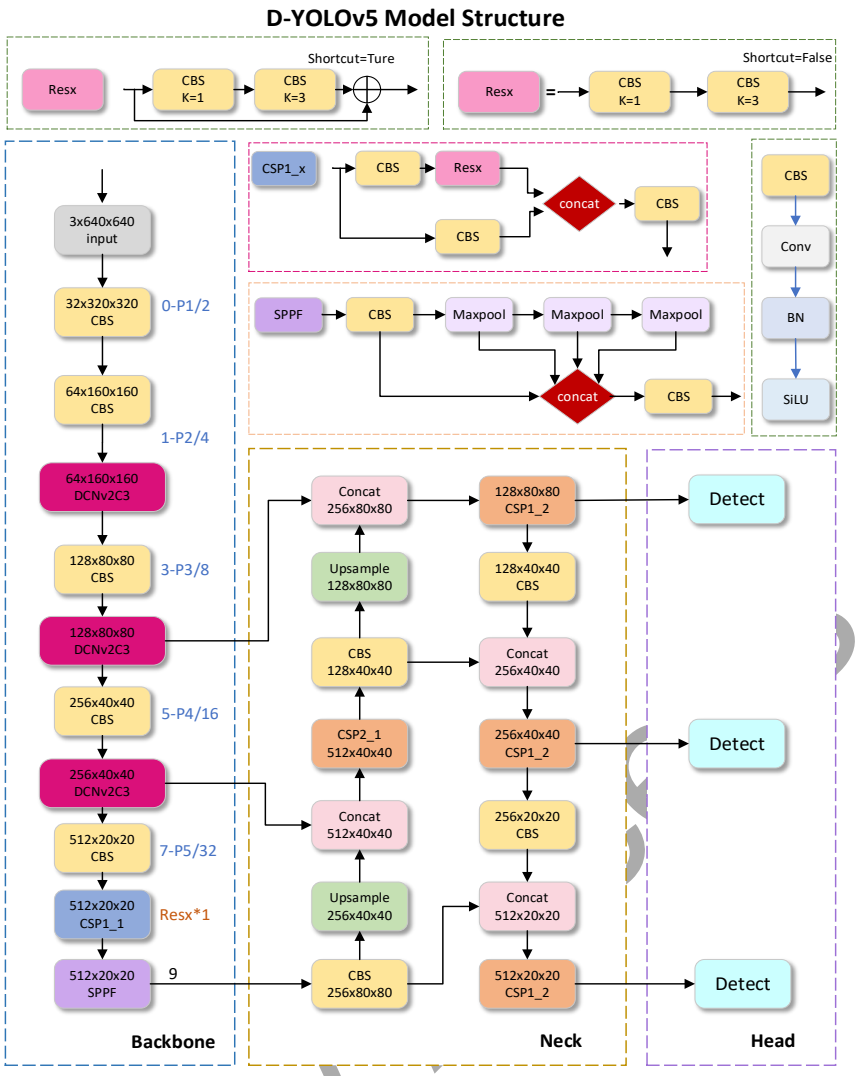


Fig. 2. The proposed DCN-YOLOv5s model for packaging detection.

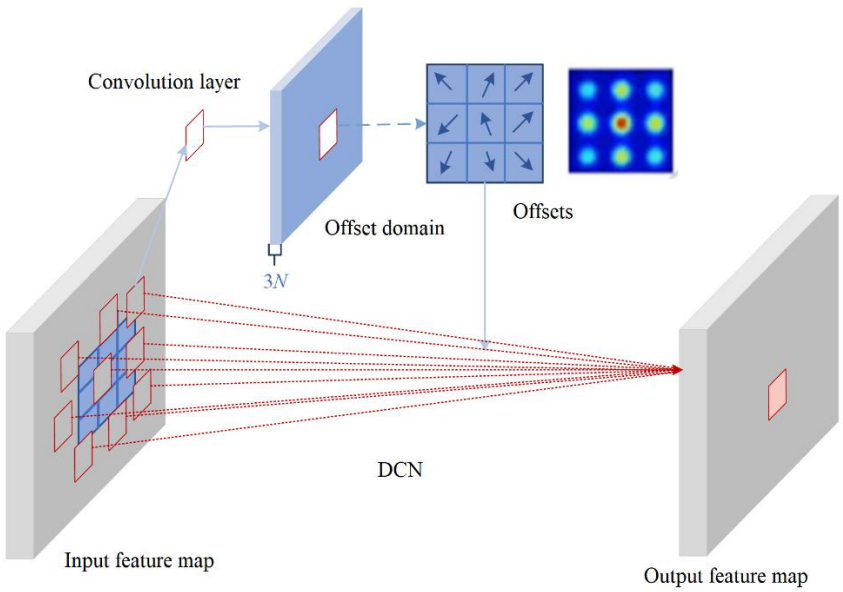


Fig. 3. Structure of DCN v2.

B. *Model Enhancement*

In object detection tasks, a common problem is that a single object can be detected by multiple bounding boxes, and the confidence scores of these boxes may all exceed the set threshold, leading to a lot of redundant detections. Non-maximum suppression (NMS) is used to address this issue. The NMS first sorts the bounding boxes by their confidence scores and selects the box with the highest score as a reference. Then, it calculates the Intersection over Union (IoU) between the reference box and the other boxes. If the

IoU exceeds a preset threshold, the boxes are considered to represent the same object, and the remaining boxes are removed. The NMS algorithm formula is as follows

$$S_i = \{b_i | Score(b_i) > \text{threshold} \forall b_i \in boxes\}, \quad (1)$$

where $Score(b_i)$ represents the score of a candidate b_i and *boxes* are a set of all candidate boxes to be processed.

Complete Intersection over Union (CIoU) further optimizes the evaluation of bounding boxes in object detection. It calculates the ratio of the distance between the centers of the ground truth and predicted boxes to the diagonal distance, taking into account three geometric factors (i.e., the overlap area, center distance, and aspect ratio). This comprehensive consideration reduces the loss value, makes the bounding box regression more accurate, and speeds up the model convergence.

C. Model Improvement

To further improve the detection accuracy in the YOLOv5s network model, the DCN v2-C3 module was used to replace the C3 module in the deep learning model. The DCN is able to address the issue that defects on packaging surfaces present similar and indistinct features by enhancing model adaptability and ability to extract high-level semantic information through modulation mechanisms. The improved DCN-YOLOv5s network is illustrated in Fig. 2.

In the images of the packaging films, the defects often contain similar features. The DCN v2, by stacking multiple deformable convolutional layers, processes the ability to learn offsets, which is crucial to improve the spatial transformation capability of the deep learning model [17]. The inclusion of the DCN v2 structure is shown in Fig. 3.

Figure 4 compares the performance of standard and deformable convolution operations using real-world production with a carton packaging defect. There are significant differences between the sampling processes of standard convolution and deformable convolution. From the perspective of the receptive field, the DCN exhibits a relatively larger sampling range when handling the production packaging defects. The generated larger sampling range not only covers the target packaging film defect regions but also incorporates a substantial amount of irrelevant background information; in contrast, the standard convolution performs strict regular sampling but does not infer any background information. As a result, if the defect exhibits features similar to the background, the standard convolution may not extract complete feature information for defect recognition.

In Fig. 4, a prominent advantage of the DCN lies in its ability to exercise finer control over the sampling process. It can adjust the sampling range over a wider range of feature levels. Using the displacement information obtained from the feature map of the preceding layer, the DCN can dynamically modify the offsets of the sampling points, thereby effectively reducing the attention to irrelevant information within the target while concentrating more on the packaging defect regions that substantially contribute to network performance. This mechanism not only enhances the specificity of feature extraction, but also strengthens the model robustness against complex backgrounds, further improving detection accuracy and efficiency.

Moreover, DCN v2 introduces a new modulation mechanism designed specifically to handle geometric changes in the target [18]. This mechanism takes into account factors like scale, pose, and texture in the target geometry, not only applying spatial offsets to the input feature maps but also dynamically adjusting the weights of inputs at each position. With this modulation mechanism, the DCN v2 can dynamically determine weights based on the target features and assess how much each sampling point contributes to the feature extraction. At the same time, this mechanism adds more flexibility, allowing inputs from irrelevant areas to be ignored when the weight is zero, effectively solving the problem of irrelevant background interference that arises in traditional methods.

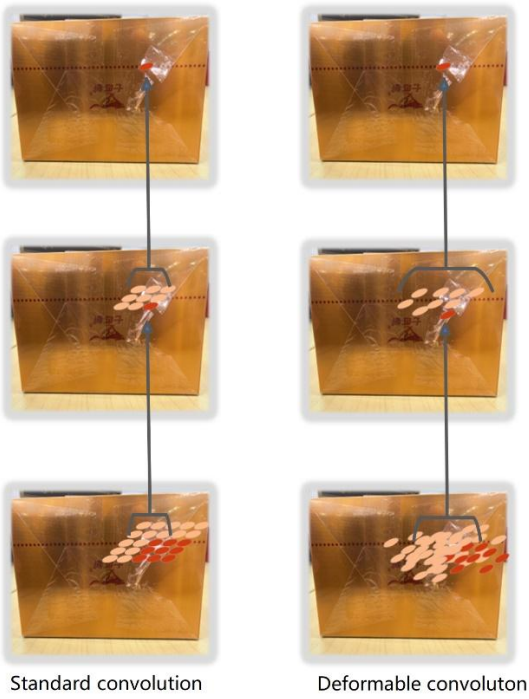


Fig. 4. Comparison of standard and deformable convolution operations.

IV. EXPERIMENTAL EVALUATION

A. Experimental Settings

The experimental environment and training parameters are shown in Table II. During deep model training, the initial learning rate is set at 0.0001, the cyclical learning rate is 0.01, the momentum is adjusted to 0.937, and the weight decay factor is 0.0005. After the initial training, the learning rate is adaptively adjusted according to the batch size.

TABLE II. EXPERIMENTAL SETTINGS.

Item	Setting
Operating System	Windows 11
Graphics Card	NVIDIA GeForce RTX 3060 GPU
Programming Language	Python 3.9
Deep Learning Framework	Pytorch 2.0.1
Development Environment	CUDA 11.8
Epochs	100
Batch Size	16
Input Image Size	640-by-640

B. Experimental Data

The quality of the dataset is an important factor in ensuring the training results of deep learning models. Since there is a

Early Ac

lack of large-scale datasets for the detection of production packaging defects on the Internet, we created our own dataset called Packaging-Defect, as shown in Table III. A total of 120 images of field data on a production line with different packaging defects were collected using front and back light.

TABLE III. EXPERIMENTAL DATA.

Condition	Label	Image number
Qualified Products	Good	10
Misaligned pull tab	Defect-1	40
Missing transparent film	Defect-2	10
Film damage	Defect-3	20
Folded corners	Defect-4	40

Using the LabelImg annotation tool, rectangular annotations were made on the four types of packaging defects in the original images to generate the YOLO format txt annotation files, which facilitates subsequent model training and testing. The annotated data are randomly split into training, validation, and test sets in an 8:1:1 ratio.

Based on the 120 original images, the data augmentation is done using vertical flipping (upside down), horizontal flipping (mirroring), center cropping, changing brightness, changing contrast, changing saturation, adding Gaussian noise, salt noise, and pepper noise [19]. The data augmentation is illustrated in Fig. 5.

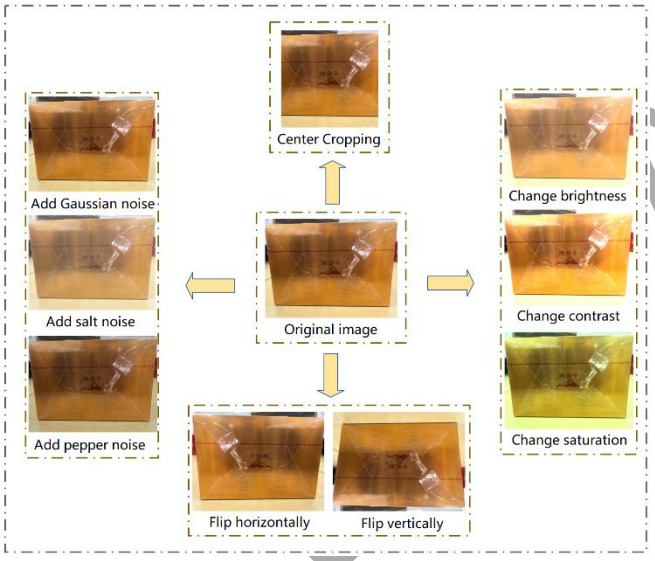


Fig. 5. Data augmentation.

After data augmentation, a total of 1,200 images are produced with four defect categories; and 960 images are used for training, 120 for validation, and 120 for testing. By using data augmentation, the feature data of the carton packaging defects in the Packaging-Defect dataset is expanded, which significantly increases the variety of detection data and boosts the deep model ability and robustness for targets recognition.

To effectively validate the effectiveness of the proposed DCN-YOLOv5s model, six indicators were selected for evaluation, including the number of model parameters (Parameters), computational complexity (Flops), precision (P), recall (R), and mean average precision (mAP@0.5, mAP@0.5-0.95) [20]. The calculation of these evaluation metrics is expressed as follows:

$$P = TP / (TP + FP), \tag{2}$$

$$R = TP / (TP + FN), \tag{3}$$

$$AP = \int_0^1 P(R)dR, \tag{4}$$

$$mAP = \frac{\sum_1^n (AP)}{n}, \tag{5}$$

where *TP* refers to the number of true positive detection boxes whose IoU with the ground truth box is greater than the threshold; *FP* refers to the number of false positive detection boxes; *FN* refers to the number of false negatives where the ground truth box was not detected; and *n* is the number of categories in the dataset. The average precision at an IoU threshold of 0.5 (mAP@0.5) and the average precision over an IoU threshold range of 0.5 to 0.95 (mAP @0.5:0.95) are commonly used as metrics to validate the effectiveness of improvement measures.

C. Analysis Results and Discussions

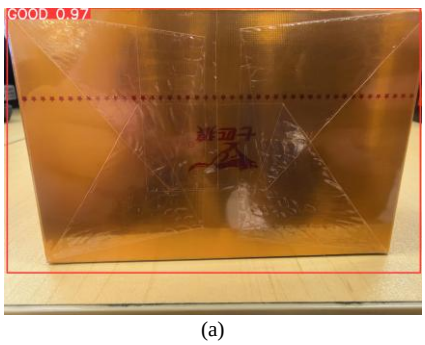
To evaluate the impact of each module on the YOLOv5s detection performance, the ablation experiments of the proposed method were conducted on the self-constructed Packaging-Defect dataset. The effects of the improvement on the model performance are presented in Table IV.

TABLE IV. ABLATION EXPERIMENT RESULTS.

Method	Parameters	Flops	*P*	*R*	mAP @0.5	mAP @0.5:0.95
YOLOv5s	6.68M	15.8G	0.981	0.965	0.982	0.966
DCN-YOLOv5s	6.74M	13.5G	0.996	0.990	0.995	0.993

As can be seen in Table IV, after introducing the DCN into the YOLOv5s model to replace the original C3 module, deeper feature information of the defect targets can be obtained. The number of model parameters has increased compared to the original YOLOv5s model, while the floating-point computation has decreased slightly and the detection accuracy has improved. The DCN-YOLOv5s model achieves an mAP@0.5 of 99.5 % and an mAP@0.5:0.95 of 99.3 %, representing increases of 1.3 % and 2.7 %, respectively, compared to the original YOLOv5s model. Considering the data in Table IV, these results validate that the proposed method is effective in achieving lightweight, fast, and high-precision detection for the carton packaging defects.

Then, the proposed DCN was applied to the testing images for packaging defect detection. To intuitively observe the detection differences between the DCN-YOLOv5s and YOLOv5s models, both were tested using the test set, with some detection results shown in Fig. 6.



(a)

Access

(b) (c) (d) (e)

Fig. 6. Experimental detection precision: (a) Detection P = 0.97 for Good condition; (b) Detection P = 0.96 for Defect-1 condition; (c) Detection P = 0.96 for Defect-2 condition; (d) Detection P = 0.96 for Defect-3 condition; (e) Detection P = 0.97 for Defect-4 condition.

As shown in Fig. 6, the actual labels of the targets in Figs. 6(a)–6(d) are, respectively, qualified products, defect-1, defect-2, defect-3, and defect-4. The proposed DCN-YOLOv5s model achieved detection accuracies of 0.97, 0.96, 0.96, 0.96, and 0.96, indicating high detection precision.

To highlight the performance of the proposed method, comparative experiments were conducted using the currently popular object detection algorithms, including Faster R-CNN, YOLOv3, and YOLOv5s. The comparison results are listed in Table V.

It can be seen from Table V that the proposed method effectively achieves the model lightweighting, with a parameter size of 6.74 M, only slightly less than YOLOv5s, but its computation Flops is the smallest 13.5 G, indicating a significant advantage over other models. It can be observed that the precision (P) and recall (R) values of the proposed method are higher than those of other models, and its mAP@0.5 surpasses the Faster R-CNN, YOLOv3, and YOLOv5s by 5.3 %, 7.8 %, and 1.3 %, respectively. The mAP@0.5:0.95 of the proposed method also exceeds the Faster R-CNN, YOLOv3, and YOLOv5s by 8.8 %, 11.9 %, and 2.7 %, indicating a substantial advantage in detection performance.

TABLE V. COMPARISON RESULTS.

Model	**Parameters**	**Flops**	***P***	***R***	**mAP@0.5**	**mAP@0.5:0.95**
Faster R-CNN	128.01 M	170.4 G	0.906	0.934	0.942	0.905
YOLOv3	61.94 M	66.0 G	0.876	0.909	0.917	0.874
YOLOv5s	6.68 M	15.8 G	0.981	0.965	0.982	0.966
DCN-YOLOv5s	6.74 M	13.5 G	0.996	0.990	0.995	0.993

Overall, the proposed method provides a good balance between computational efficiency and detection accuracy. We installed the well-trained DCN-YOLOv5s model in the online monitoring system of a production packaging line. Online detection results demonstrate that the installed deep learning mode meets the practical requirement for the calculation speed and maintains a detection accuracy of 95.15 % for carton packaging defects.

V. CONCLUSIONS

This work addresses the problems in the current manual inspection of production packaging defects, such as large subjective errors, visual fatigue, and low accuracy. A deep learning-based method is proposed to improve detection accuracy and calculation speed. To tackle the issue of insufficient feature extraction for packaging defects with similar shapes and complex textures, an adaptive information extraction module using the DCN v2-C3 is introduced into the YOLOv5s network. By stacking deformable convolutional layers, feature extraction sampling can be controlled to enhance the generalization ability of the detection model. The key results are listed as follows.

1. The detection performance for production packaging defects is better than that of the original YOLOv5s model in the laboratory experiment test, with mAP@0.5 increased from 0.982 to 0.995, and mAP@0.95 increased from 0.966 to 0.993.
2. The present online detection system is evaluated using a real-world production line and the online detection rate is 95.15 %.

In the future, the network structure can be further optimized, focusing on improving detection accuracy and building high-quality datasets to achieve better detection performance. In addition, the well-trained detection model will be installed in other production packaging lines to evaluate its generalization ability.

Early Access

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest. Longyan Tobacco Industry Co., Ltd. has only supported the basic research activities in this work and will not use the technique developed for any commercial products.

REFERENCES

[1] X. Chen and X. Li, "A design of a visual detector for cigarette pack appearance based on the original device", *Instrumentation and Equipments*, vol. 11, no. 2, pp. 95–103, 2023. DOI: 10.12677/iae.2023.112013.

[2] W. Wu, A. Liu, and Y. Wang, "Study on visual detection methods for defects in transparent cigarette packaging", *Journal of Yunnan University*, vol. 31, pp. 116–119, 2009.

[3] Y. Zhao, Q. Zheng, D. Xu, L. Zhang, "Design of software system for cigarette appearance inspection", *Machinery Manufacturing and Automation*, vol. 41, no. 2, pp. 132–134, 2012. DOI: 10.3969/j.issn.1671-5276.2012.02.044.

[4] X. Yan, "Cigarette anomaly detection algorithm based on visual perception features", *Tobacco Science and Technology*, vol. 49, no. 1, pp. 78–83, 2016. DOI: 10.16135/j.issn1002-0861.20160112.

[5] P. Cai, W. Hua, X. Jiang, Z. Zhu, L. Zhong, L. Chen, *et al.*, "Design of a BV packaging machine cigarette appearance quality inspection device", *Packaging Engineering*, vol. 39, no. 23, pp. 143–150, 2018. DOI: 10.19554/j.cnki.1001-3563.2018.23.025.

[6] B. Song, Y. Wang, and L.-P. Lou, "SSD-based carton packaging quality defect detection system for the logistics supply chain", *Ecological Chemistry and Engineering S*, vol. 30, no. 1, pp. 117–123, 2023. DOI: 10.2478/eces-2023-0011.

[7] Y. Wu and Y. Lu, "An intelligent machine vision system for detecting surface defects on packing boxes based on support vector machine", *Measurement and Control*, vol. 52, nos. 7–8, pp. 1102–1110, 2019. DOI: 10.1177/0020294019858175.

[8] N. Sun, "Research on the online detection method of cigarette packaging appearance quality", M.S. thesis, Kunming University of Science and Technology, Kunming, China, 2020.

[9] X. Yang, M. Han, H. Tang, Q. Li, and X. Luo, "Detecting defects with support vector machine in logistics packaging boxes for edge computing", *IEEE Access*, vol. 8, pp. 64002–64010, 2020. DOI: 10.1109/ACCESS.2020.2984539.

[10] R. Chen, W. Zhang, Z. Guo, W. Wu, "Cigarette pack appearance defect detection system based on YOLOv4 convolutional neural network", *Mechanical Design and Manufacturing Engineering*, vol. 52, no. 11, pp. 86–90, 2023. DOI: 10.3969/j.issn.2095-509X.2023.11.018.

[11] W. Li *et al.*, "Online detection system for adhesive quality of cigarettes carton based on improved GhostNet-YOLOv4", in *Proc. of 7th International Conference on Vision, Image and Signal Processing (ICVISP 2023)*, 2023, pp. 1–4. DOI: 10.1049/icp.2023.3275.

[12] Z. Liang *et al.*, "Intelligent recognition and handling of carton packing straps using 3D laser vision and deep learning", in *Proc. of 2025 International Conference on Industrial Automation and Robotics*, 2025, pp. 85–90. DOI: 10.1145/3778886.3778899.

[13] J. Pan, J. Yang, Y. Liu, and Y. Lv, "Oriented R-CNN with disentangled representations for product packaging detection", *IEEE Photonics Journal*, vol. 16, no. 5, pp. 1–11, 2024. DOI: 10.1109/JPHOT.2024.3450295.

[14] N. Susanti, C. E. Widodo, and A. Setiawan, "A systematic review of AI-based packaging defect detection to strengthen digital industrial resilience and sustainability", in *Proc. of 7th International Conference on Informatics, Multimedia, Cyber and Information System, ICIMCIS 2025*, 2025, pp. 378–383. DOI: 10.1109/ICIMCIS68501.2025.11326917.

[15] J. Glenn, "Ultralytics YOLOv5", Software License AGPL-3.0, 2020. [Online]. Available: https://github.com/ultralytics/yolov5

[16] Q. Wu, X. Chen, C. Chen, F. Zhang, S. Wang, H. Zhao, "A method for identifying mushroom log contamination based on improved YOLOv5s", *Chinese Journal of Agricultural Mechanization*, vol. 46, pp. 217–223, 2025. DOI: 10.13733/j.jcam.issn.2095-5553.2025.02.032.

[17] X. Zhu, H. Hu, S. Lin, and J. Dai, "Deformable ConvNets V2: More deformable, better results", in *Proc. of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 9300–9308. DOI: DOI: 10.1109/CVPR.2019.00953.

[18] W. Liu, D. Liu, and L. Wang, "A review of deformable convolutional networks", *Computer Science and Exploration*, vol. 17, pp. 1549–1564, 2023. DOI: 10.3778/j.issn.1673-9418.2209076.

[19] B. Yan, P. Fan, X. Lei, Z. Liu, and F. Yang, "A real-time apple targets detection method for picking robot based on improved YOLOv5", *Remote Sensing*, vol. 13, no. 9, p. 1619, 2021. DOI: 10.3390/rs13091619.

[20] Z. Gui, J. Chen, Y. Li, Z. Chen, C. Wu, and C. Dong, "A lightweight tea bud detection model based on Yolov5", *Computers and Electronics in Agriculture*, vol. 205, art. 107636, 2023. DOI: 10.1016/j.compag.2023.107636.

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution 4.0 (CC BY 4.0) license (http://creativecommons.org/licenses/by/4.0/).

Early Access