






## Enhanced Express Package Trademark Recognition via a Novel PTD-YOLO Algorithm

Yanghua Gao<sup>1\*</sup>, Weidong Lou<sup>1</sup>, Hailiang Lu<sup>1</sup>, Li Zhang<sup>2</sup>, Yaliang Wang<sup>2</sup>

<sup>1</sup> Information Center, China Tobacco Zhejiang Industrial Co., Ltd., Hangzhou 310008, China

<sup>2</sup> College of Mechanical Engineering, Zhejiang University of Technology, Hangzhou 310032, China

Corresponding Author Email: [yhgao@zjtobacco.com](mailto:yhgao@zjtobacco.com)

<https://doi.org/10.18280/ts.400324>

**Received:** 8 March 2023

**Accepted:** 25 May 2023

### Keywords:

packing detection, trademark, deep learning, YOLO

### ABSTRACT

The rise in prominence of the logistics industry necessitates a boost in its efficiency. A notable hurdle to this lies in the classification of goods, based on the unique trademark of express packages, a problem with a direct bearing on delivery efficiency. Traditional methodologies for inspecting packaging appearances struggle with accuracy in recognizing a variety of scales, necessitating the use of multiple detection systems. Additionally, they fail in accurately ascertaining the precise location and size of the express packaging trademark. To rectify this, the study presents the development and application of a detection technique christened PTD-YOLO (Packing Trademark Detection algorithm based on YOLO, PTD-YOLO). This technique bolsters the YOLO v5 algorithm through improvements in three key areas. The first is the restructuring of the FSRP (Focus module with Structural Re-Parameterization) module, aimed at enhancing pre-backbone features. The second involves the integration of a novel prediction head, designed to bolster the ability of PTD-YOLO in detecting smaller-scale targets. Lastly, an attention mechanism has been incorporated within the head part, to better distinguish relevant features of detected objects. The performance of the PTD-YOLO has been validated via rigorous ablation and comparative experiments, proving its effectiveness and reliability.

## 1. INTRODUCTION

The labor-intensive nature of packaging classification, combined with the high workload and limited inspection time, renders the current methodology for packaging appearance inspection suboptimal [1]. At present, manual visual inspection forms the backbone of this process, leading to a myriad of challenges, including the potential for subjective results, inconsistencies, oversights, misidentification, and arbitrary classifications. This manual approach hampers the efficiency of detection, a concern heightened by the escalating demand for product customization. Moreover, the precision of detection holds significant implications for the evaluation of both product quality and pricing.

As production continues to evolve, there has been an amplification in the demand for a diverse range of packaging types and sizes. This diversification further compounds the difficulty of manually identifying and classifying the packaging surface. In response to these challenges, the development of an intelligent express sorting and handling system has become a pressing need, with accurate packaging trademark identification forming a critical component of this endeavor.

For automated packaging trademark inspection, innovative appearance inspection methods need to be formulated. Contemporary inspection methods are based on algorithms that identify specific areas in camera images for detection, and target detection algorithms are integral to these methods [2]. Over time, a series of robust target detection algorithms have been developed, each producing commendable recognition results. The progression of these algorithms can be broadly categorized into three stages.

The first stage was marked by object detection algorithms predominantly focusing on precise image segmentation. Angelova and Zhu [3] proposed a method that detects low-level regions for full-object segmentation, which distinguishes various types of objects through fine-grained recognition. Gould et al. [4] highlighted the intricate relationship between object detection and multi-class image segmentation, introducing a hierarchical region-based method that linked these tasks based on a coherent probabilistic model. Moreover, Carreira and Sminchisescu [5] presented a framework that formed a regular image grid based on Constrained Parametric Min-Cut problems (CPMC), thereby maximizing the category of segmentation. However, these algorithms, being task-specific, lacked the versatility required for wider application.

Following this, the focus shifted to the application of inspection technology based on machine vision. Paliwal et al. [6] integrated morphological, color, and texture features to distinguish five grain types through a neural network. This method achieved approximately 90% accuracy in some grains, despite variations in size and appearance. In another study, Meng et al. [7] designed a machine vision system to assist with mechanical inter-row weeding. They used the hue component for image processing and computed the H component for generating grayscale images to mitigate the adverse impact of light. This technique enabled the system to determine the endpoint of the images, thereby tracking the edge of weeds effectively. Additionally, Qureshi et al. [8] considered two critical fruit features, texture and shape, and utilized support vector machines as a classifier for their experiments. The method demonstrated high precision in various environments.

Despite these advancements, the field continues to grapple with several challenges, not least due to the increasing

diversification of the manufacturing landscape. Consequently, the use of Convolutional Neural Networks (CNN)-based object detection algorithms, renowned for their strong representational power and high versatility, has begun to supplant conventional methods.

These CNN-based defect detection networks can be subdivided into two structures: one-stage and two-stage algorithms [9]. A series of two-stage algorithms, including Fast R-CNN [10], Mask R-CNN [11], and Faster R-CNN [12], were proposed by Girshick et al. [10-12]. While these algorithms are known for their accuracy, their relatively slow speed makes them less suited for packaging trademark detection. To address this, one-stage algorithms have been developed which transform the classification problem into a regression problem. These algorithms directly predict position information, leading to an improvement in speed [13].

A diverse array of strategies have been employed in prior research to improve the accuracy and efficiency of object detection models, though these approaches often remain ill-suited for the unique challenges presented by carton damage detection scenarios. Carton damage detection scenarios are characterized by low color contrast and variable damage sizes, making it susceptible to errors.

A compelling contribution in the domain of multi-scale targets in ship detection was proposed by Guo et al. [14]. A novel model was proposed that ameliorates the backbone module and enhances spatial adaptability via a novel loss function. Empirical results revealed that the proposed model exhibits superior multi-scale detection capability and mobile platform portability.

In a different approach, Han et al. [15] devised a lightweight model and improved the network structure employing the K-means++ clustering algorithm. The modified model has shown an enhancement in the detection performance of small-size targets.

Zhang and Wen [16] directed their efforts towards mitigating the high maintenance cost of wind turbine blades (WTB). A surface defect detection algorithm for WTB, dubbed SOD-YOLO, was developed. A fusion of the K-means algorithm and attention mechanism to YOLO v5 was implemented following the collection of a WTB dataset. This model resulted in a 7.82% mAP and 28.3% increase in detection speed.

Simultaneously, Qi et al. [17] made improvements to the YOLO network structure for tea chrysanthemum detection during different stages. The methodology combines data augmentation, transfer learning, and other methods for the flowering stage detection. With a mAP of 92.49% on the test set, it demonstrated a remarkable performance and can be trained on a single GPU.

Aiming at avian detection, Siriani et al. [18] proposed an enhanced YOLO v4 model to locate birds. The incorporation of the Kalman filter allowed tracking of chickens in low light environments. This algorithm achieved a high-performance rate in bird detection even in high noise images.

Despite these advancements, the need for an efficient algorithm tailored to packaging appearance inspection is apparent. In response to this challenge, a package appearance damage target detection algorithm, known as PTD-YOLO (Packing Trademark Detection algorithm based on YOLO), is proposed. This method reconstructs the Focus module, introduces an additional detection scale, and improves the prediction header structure, making it more compatible with the packaging appearance inspection task. Compared with

existing target detection algorithms, PTD-YOLO achieved high accuracy [14-18].

Three crucial contributions are put forward:

- (1) Considering the limited data of packaging appearance, model reparameterization technology was introduced into the Focus module to enhance the mining of existing data features. A deeper network structure was designed through multiple branches.
- (2) To accommodate various sizes of trademark images, a four head prediction structure was constructed. This new prediction head collaborates with the original three heads to meet the detection needs of drastically changing target scales.
- (3) Based on the reconstructed Focus and four-head prediction structure, the PTD-YOLO algorithm was proposed. By introducing an attention mechanism in the head before the output of object detection, the box reflection problem was overcome. This method assists the algorithm in targeting high-latitude features more related to the target trademark.

While these contributions mark a step forward in packaging appearance inspection, further investigations are warranted to fully realize the potential of these techniques in real-world applications. The expansion of this domain of research may provide further improvements and efficiencies in the field, facilitating enhanced packaging quality, and the delivery of superior products to consumers.

## 2. PTD-YOLO

The YOLO algorithm, widely recognized as an object detection technique, is designed to deliver both the class probability and positional coordinates of objects. Among its various iterations, YOLO v5 has been a prominent variant. Comprising four modules—input, backbone, neck, and prediction head—YOLO v5 has displayed exceptional speed and precision, delivering commendable performance on open-source databases. However, its efficacy for detecting and counting trademarks on packaging images in industrial environments has been found wanting. The challenge arises from the small target nature of trademark detection on a carton surface, coupled with the multitude of detection tasks. Additionally, issues like trademark distortion and package reflection due to specific shooting angles adversely affect its detection speed.

To address these limitations, enhancements have been proposed to the Packing Trademark Detection algorithm based on YOLO (PTD-YOLO) (as shown in Figure 1). This paper outlines a strategy to augment the performance of PTD-YOLO in three distinct ways, which is anticipated to be effective in detecting targets on express packaging. These modifications have been suggested with the objective of enhancing its detection capabilities for packaging targets.

Initially, the Focus module has been restructured into the Focus module with Structural Re-Parameterization (FSRP). This innovative approach allows the creation of new slices from multiple slices. Following this, a small target detection layer was incorporated to detect trademarks of varying sizes, thus expanding the detection range and bolstering the image feature extraction capability. Finally, an attention mechanism has been introduced in the Prediction Head module to ensure more accurate brand positioning. This revised algorithm has undergone training through transfer learning.

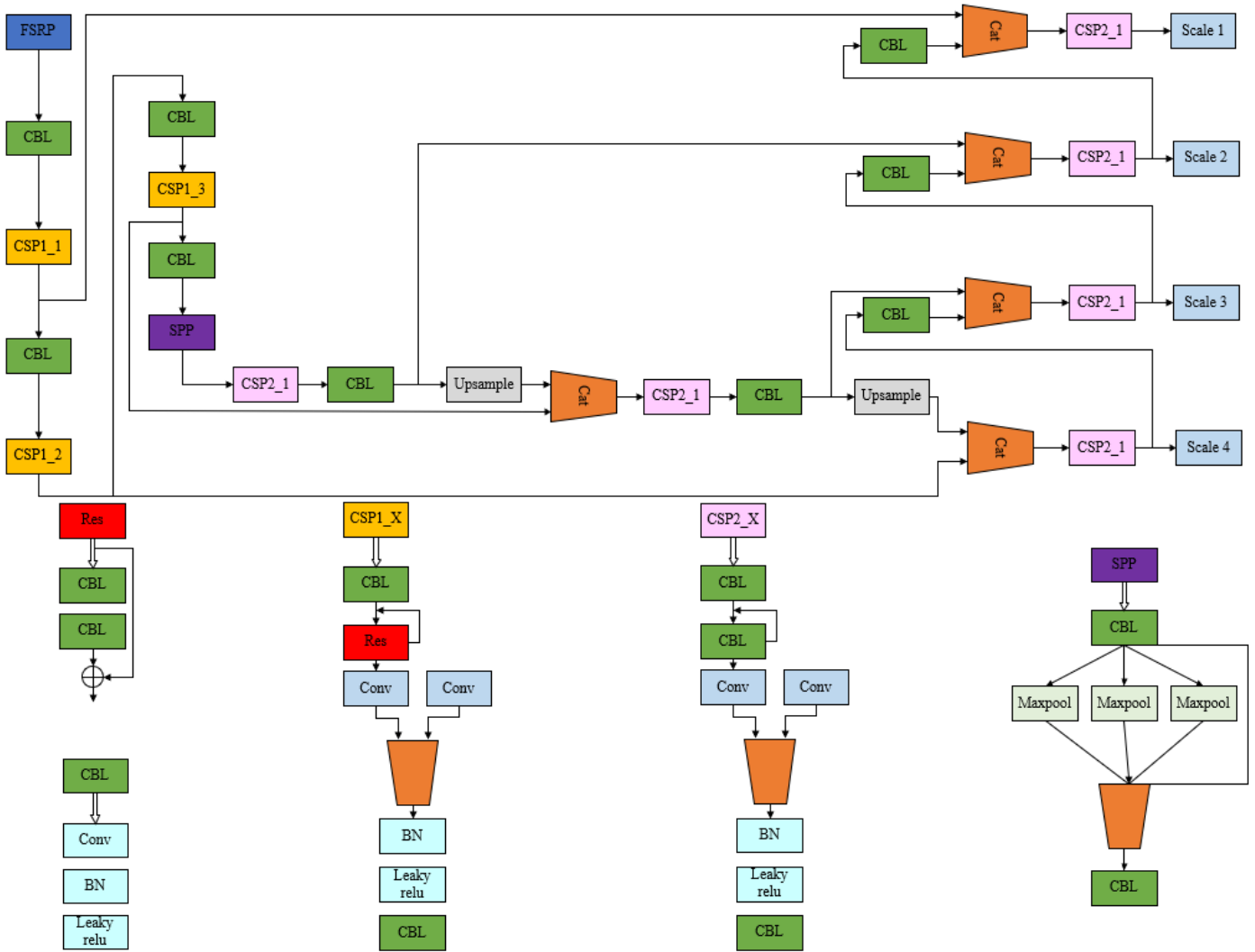


Figure 1. Structure of PTD-YOLO

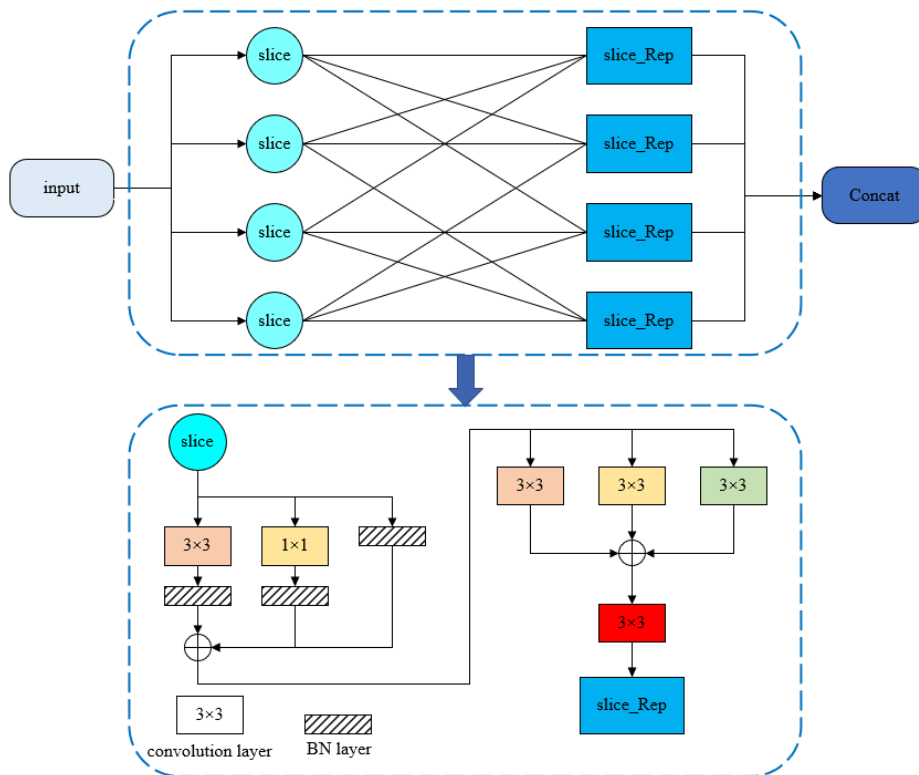


Figure 2. Structure of FSRP module

## 2.1 Focus module with Structural Re-Parameterization

In YOLO v5, the multi-layer structure effectively segregates the target area of the image. The original approach involved the entry of images into the network through the Focus module, whereby a slice operation is conducted on the image of each channel. However, for small target detection, direct subsampling tends to omit small target information from the original image, leading to sub-optimal feature extraction of small targets, which could potentially result in missed detection.

To tackle this issue, the FSRP module was developed, based on the Focus module (as shown in Figure 2). The underlying principle of FSRP is to amalgamate the four slices in focus in pairs according to the channel, to form four new slices, which are subsequently merged. The objective of FSRP is to enhance early feature selection via two slices. Taking cues from ResNet, the FSRP module leverages a deeper network structure with multiple branches. However, such an approach also induces lower network parallelism and higher memory consumption, potentially hindering the future deployment of PTD-YOLO on terminal devices with limited computational capabilities.

Therefore, in an attempt to resolve these issues, the FSRP module integrates the RepVGG model reparameterization technique. This strategy amalgamates the advantages of multi-branch network structures with the high parallelism and low memory access of straight cylinder network structures. RepVGG distinguishes between the training and inference networks, and its multi-branch network structure lacks an activation function during training.

In the FSRP module, the RepVGG structure is recombined through three steps: the fusion of convolution (Conv) and batch normalization (BN), unification of convolution kernels of different sizes into the same size, and fusion of convolution kernels of uniform size into a single convolution kernel.

Finally, the four reparameterized slices are concatenated through fusion and input into the backbone.

By applying these principles, the study endeavors to provide a more robust algorithm for detection tasks, specifically targeting industrial packaging scenarios. It is envisaged that this will offer a significant boost to detection speed and efficiency, opening new avenues in express packaging management. Further research in this direction could potentially expand the applications of this algorithm, contributing to the broader field of object detection and classification.

## 2.2 Prediction head with four scales

In the conventional single feature scale approach, features are directly detected from the original image, leading to the extraction of only the most conspicuous features and the potential loss of high-level components. YOLO v5, with its three feature scales, aims to deepen high-level feature extraction, but is less suitable for the detection of small-scale trademarks on express packaging.

The Backbone network of YOLO v5 functions primarily as a classification network and does not undertake localization tasks. Those are delegated to the Head network, which is responsible for detecting tasks on feature maps extracted by the Backbone. It can be observed that while the original YOLO v5 algorithm can detect at three scales, providing relatively balanced results, it lacks effectiveness in small target detection.

Given the large target scale of the packaging trademarks being analyzed in this context, minimal target detection is a crucial requirement. As such, an addition has been proposed to this system in the form of a Prediction Head with Four Scales (PHFS) (as shown in Figure 3). The PHFS is designed to work in tandem with the original three heads and is anticipated to effectively match the detection requirements of the drastically varied target scale.

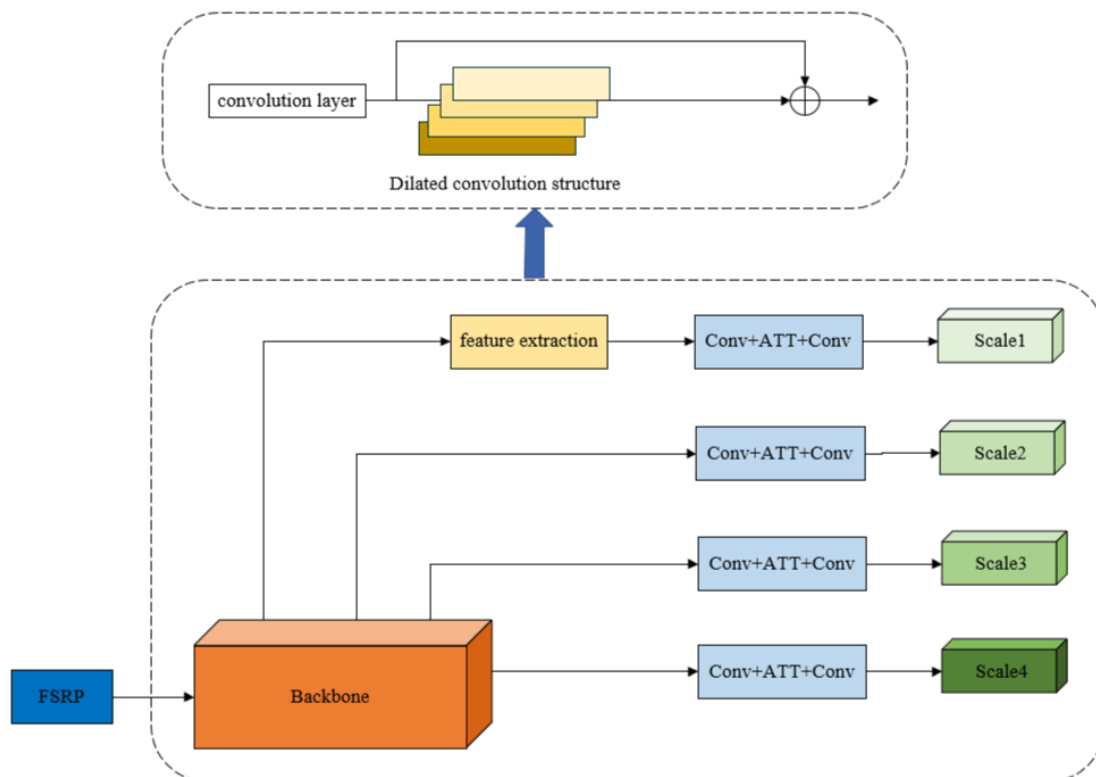


Figure 3. Four head structure

An additional feature scale has been integrated to enhance the recognition of subtle changes at the image edge. Furthermore, feature fusion is performed on the down-sampled image, mitigating information loss due to resolution reduction. The fused feature scale, therefore, contains information from previous scales, thereby improving detection effectiveness. Specifically, an upsampling and tensor stitching has been introduced to the Neck part of the YOLO v5 algorithm. This process involves the extraction of an extra layer of features and the performance of an additional tensor stitching, expanding the amount of data in the feature dimension and enhancing model accuracy for small target detection.

Higher-resolution images are indispensable for the detection of smaller targets, necessitating the extraction of deeper features. The feature extraction structure of the newly-added prediction head, therefore, introduces multi-scale dilation convolution. This dilated convolutional layer equips the lightweight CNN to extract deeper features and to more effectively expand the algorithm’s receptive field.

The feature extraction structure of the newly-added prediction header is comprised of a multi-branch convolutional layer and a dilation convolutional layer (as shown in Figure 4). Each branch of the feature extraction structure is connected to a dilation convolution operation of varying size after a normal convolution operation, to obtain higher-resolution features. To curb computational cost, multiple  $1 \times 1$  convolutional layers are utilized to increase network depth and width. Furthermore, a nonlinear excitation layer is appended to the  $1 \times 1$  convolutional layer to enhance network expressiveness. A residual structure is incorporated into the feature extraction structure of the new prediction head to increase feature expression, as the mere modification of the convolution kernel size proves insufficient. Lastly, the feature extraction structure of the new prediction head integrates four  $3 \times 3$  convolutional layers, enabling the model to capture deeper features.

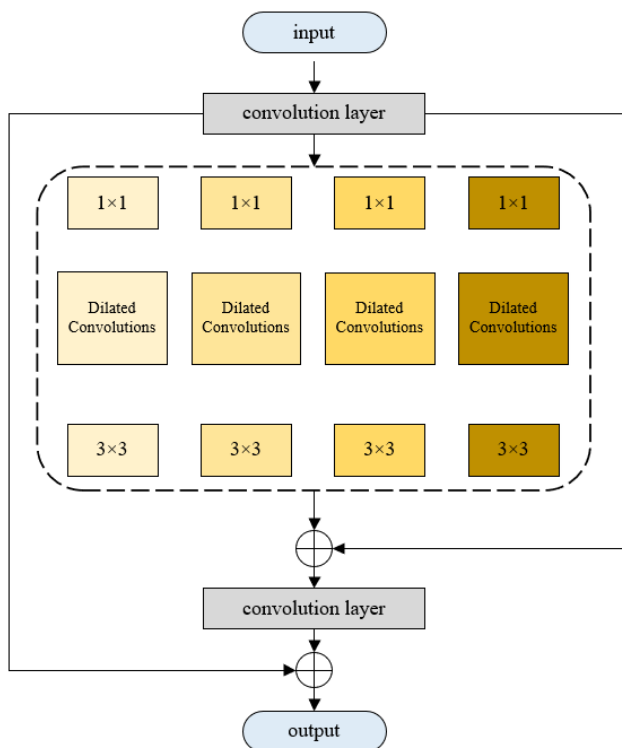


Figure 4. Structure of dilated convolution

Through the application of these principles, this study aspires to augment the capabilities of the algorithm for detection tasks, specifically in the context of industrial packaging scenarios. It is envisaged that this enhanced system will provide substantial improvements to detection speed and efficiency, heralding new possibilities in the management of express packaging. It is anticipated that further exploration in this direction could broaden the applications of this algorithm, contributing significantly to the wider field of object detection and classification.

### 2.3 Enhanced predictive head module

The prediction head serves as a critical constituent in target detection algorithms, primarily containing convolutional neural networks, as it significantly impacts detection result precision. The detection of packaging trademarks, particularly those of monochromatic nature, challenges the capture of essential high-order features. In addition, multiple stacked cartons in images present complexities in identification due to the phenomenon of local reflections.

Fortunately, the introduction of attention mechanisms has led to fresh avenues in the study of visual characteristics [19], allowing the distribution of various weights to features for subsequent tasks, thereby homing in on more pertinent features. When contrasted with traditional convolutional neural networks, visual models leveraging attention mechanisms have achieved remarkable triumphs in numerous fields.

This leads to the incorporation of the attention mechanism into the predictive header structure of PTD-YOLO, thus refining the original YOLO v5 prediction header structure, as shown in Figure 5. This innovative approach allows for the capture of comprehensive global information and nuanced contextual details, enabling precise positioning of package trademarks, even in scenarios characterized by low-discrimination or reflective surfaces.

### 2.4 Training strategy

Considering the constrained dataset for carton trademark images, a training strategy employing transfer learning is utilized during the model training process. Transfer learning [20] has been extensively applied across various sectors, enhancing training by fine-tuning parameters and successively optimizing the model.

For the present study, a pre-training phase is adopted, using a packaging image dataset sourced from the internet to procure a pre-trained model. This model is then updated and optimized through additional packaging image datasets, obtained via cameras and mobile phones, as depicted in Figure 6. The training procedure is divided into two steps:

(1): PTD-YOLO’s backbone network, CSPDarkNet-53, is trained on a packaging image set collected from the network to yield a pre-trained model.

(2): The pre-trained model is subsequently loaded into PTD-YOLO, fixing the parameters of the pre-trained backbone network, before training PTD-YOLO using the packaging image dataset captured by cameras and mobile phones. The parameters of the Neck and Head at the backbone are the only ones updated and retrained.

Images of packaging trademarks obtained from the network share fundamental features, such as color, outline, and texture, with images procured through cameras and mobile phones.

These shared attributes are considered lower-level information and reside near the input end, retaining a significant amount of this information, resulting in the features extracted by the backbone part being generalizable.

In the present study, the pre-trained model is achieved by pre-training the model with the packaging image dataset collected from the network, which boasts a larger size and greater diversity, thereby resulting in superior model performance. The weights from the pre-trained model are then

loaded into PTD-YOLO, and the parameters of the pre-trained Backbone model are frozen, subsequently reducing the number of network layers.

This approach addresses the issue of overfitting, which can stem from insufficient data, while also requiring less training data. The result is increased training efficiency and an enhancement of the model's generalization capability and robustness.

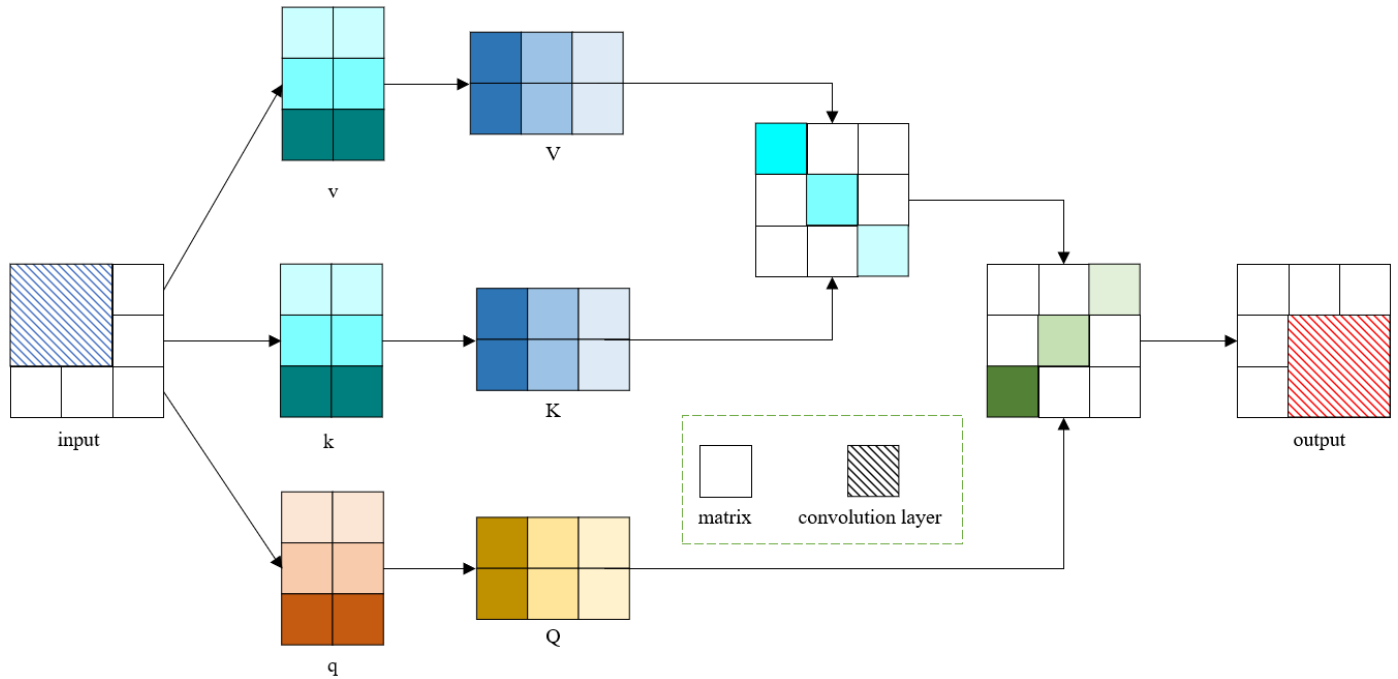


Figure 5. Improve structure of prediction head

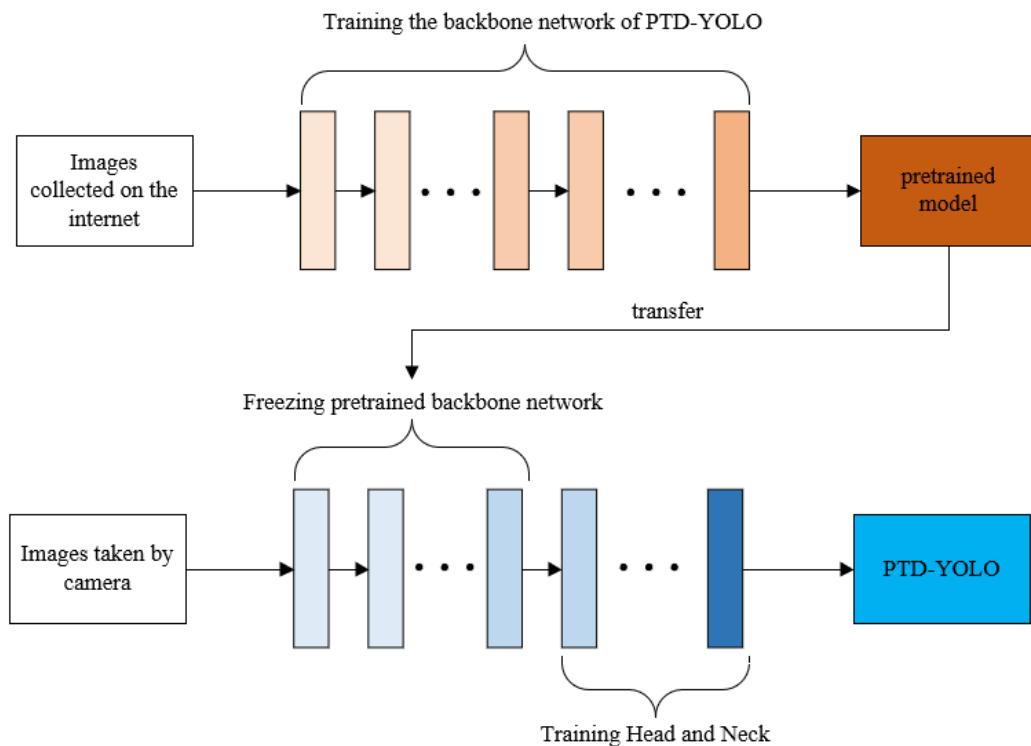


Figure 6. Training strategy of PTD-YOLO

### 3. EXPERIMENTAL PROCEDURE AND FINDINGS

#### 3.1 Parameter configuration for training

The training of the code was expedited via CUDA, executed on an RTX 3090 24G graphics card, under the Windows 11 operating system. The deep learning framework, PyTorch, in its 11.0 version was harnessed for the creation of the model. Complementarily, the computational resources included an Intel i7-12700KF CPU, paired with a 64GB memory.

#### 3.2 Indices of evaluation

The experiment was gauged on the grounds of precision P, recall R, average precision AP, and mean average precision mAP. AP was computed as the area bordered by the Precision-Recall (P-R) curve and the coordinate axis. For each of the n categories, True Positives (TP), False Positives (FP), and False Negatives (FN) were calculated, according to the defined formula as follows:

$$P = \frac{TP}{TP + FP} \quad (1)$$

$$R = \frac{TP}{TP + FN} \quad (2)$$

$$AP = \int_0^1 (P(R)dR) \quad (3)$$

$$mAP = \frac{\sum (AP)}{n} \quad (4)$$

#### 3.3 PTD-YOLO ablation experiment

The effects of technological advancements, such as FSRP, PHFS, and Attention on the recognition accuracy of PTD-YOLO were perceptively examined through ablation experiments as reflected in Table 1.

Following the introduction of FSRP, an enhancement of 0.8% was recorded in the mAP of the PTD-YOLO model. With the further application of a four-head detection structure, the mAP noted an incremental rise of 1.5%. The integration of an attention mechanism into the prediction header fostered a more effective extraction of features from the convolutional structure, culminating in a mAP ascent of 0.3%. The combined effects of these advancements resulted in a remarkable boost in detection accuracy. Particularly, FSRP+MFS exhibited a mAP improvement of 2.3%. When FSRP was combined with a prediction header equipped with an attention mechanism, a performance enhancement of 1.7% was observed, translating to a mAP of 97.7%. When the four-head structure was coupled with an attention-stimulating predictive head, a mAP improvement of 2.5% was recorded. Remarkably, the four-head structure excelled over the other two enhancements, with the first seven ablation experiments demonstrating that the four-head configuration resulted in superior mAPs (1.5%, 2.3%, and 2.5% higher than baseline, respectively). The convergence of all three enhancements exhibited the most profound effect, signifying a 3.1% mAP increase, thus, affirming the efficacy of the proposed algorithm.

**Table 1.** Comparative analysis of experimental results from various modules of PTD-YOLO

Module	mAP%
YOLO v5	95.6%
YOLO v5+FSRP	96.4
YOLO v5+PHFS	97.1
YOLO v5+Attention	95.9
YOLO v5+ FSRP+PHFS	97.9
YOLO v5+ FSRP + Attention	97.3
YOLO v5+PHFS+ Attention	98.1
YOLO v5+ FSRP+PHFS+ Attention	98.7

#### 3.4 Performance comparative experiment

In order to assess the superiority of the PTD-YOLO, it was compared with several conventional algorithms including the Two-Stage algorithms, the well-known SSD algorithm and a series of YOLO algorithms under One-Stage detectors. For the task of carton logo detection, YOLO v2, YOLO v3, YOLO v3-tiny, YOLO v4, YOLO v4-tiny, YOLO v5, and YOLO v5-tiny were chosen for comparison with PTD-YOLO. The Mean Average Precision (mAP) was employed as the evaluation index in this analysis. Identical hyperparameters were employed in the training and testing of each model. The outcomes of the detection are delineated in Table 2.

**Table 2.** Detection performance comparison among different algorithms

Algorithm	mAP%
Fast R-CNN	62.3
Faster R-CNN	70.1
YOLO v2	80.6
YOLO v3	84.2
YOLO v3-tiny	79.3
YOLO v4	90.7
YOLO v4-tiny	88.2
YOLO v5	95.6
YOLO v5-tiny	87.2
SSD	78.2
PTD-YOLO	98.7

On scrutinizing the results depicted in Table 2, it can be discerned that the Two-Stage detection algorithms, Fast R-CNN and Faster R-CNN, despite exceeding 60% in detection accuracy, fall short of meeting the research objectives and industrial standards for carton logo detection, possibly due to their inadequacy in detecting small-scale trademarks. The SSD algorithm likewise fails to attain high accuracy. Conversely, the YOLO series algorithms demonstrate substantial progress, with mAP values surpassing 80%. YOLO v4 and v5 algorithms have even attained mAP values over 90%, signaling their potential for practical implementation. Out of the eleven algorithms evaluated, PTD-YOLO achieves superior performance, marking a maximum improvement of 36.4% in detection accuracy compared to the Two-Stage algorithm. Even in comparison with the improved YOLO v2, v3, v4, and v5, this algorithm displays improvements varying from 3.1% to 18.1% in detection accuracy, suggesting a more precise estimation of trademark position and size distribution. In sum, PTD-YOLO surpasses in both detection accuracy and real-time express packaging surface logo detection, offering superior comprehensive performance and practical utility. Figure 7 showcases an exemplary application of PTD-YOLO.





**Figure 7.** Illustrative Diagram of a Practical Application of PTD-YOLO

### 3.5 Constraints of PTD-YOLO

Despite the commendable accuracy of the proposed PTD-YOLO algorithm in tasks of carton trademark recognition, thus paving the path for potential future applications, certain limitations are also observed.

(1): The introduction of the FSRP module, while enhancing the accuracy of carton trademark detection, necessitates the incorporation of a new detector to broaden the detection range and distribution. However, akin to conventional object detection models, the implementation of any new modules demands robust terminal device support to ensure successful deployment of the model in practical production tasks, thereby escalating the economic costs.

(2): Furthermore, stability of the algorithm becomes a paramount consideration for the successful application of object detection algorithms in the realm of carton trademark detection. Given that these algorithms are primarily devised and evaluated in experimental environments, performance comparisons are often conducted within limited theoretical frameworks. Hence, to effectuate the application of these object detection algorithms in real-world scenarios, their stability must be ascertained under conditions of continuous operation.

### 4. CONCLUSION

The PTD-YOLO algorithm is put forward in this work as a response to challenges within deep learning algorithms pertinent to carton appearance trademark detection. Notably, the algorithm advances the FSRP module to facilitate multiple coverage, ensuring exhaustive extraction of early features and thwarting the omission of smaller targets. With the varied size diversity of carton appearance trademarks taken into account, the algorithm unveils an improved prediction head and detection scale. Consequently, detection precision is refined across trademarks of diverse sizes, addressing the hurdles associated with low color discrimination and detection of small target packaging appearance trademarks. Experimental data validates the enhancement of detection accuracy and

detection ability for small target trademarks facilitated by PTD-YOLO.

The work also brings forth a carton appearance trademark detection model that is tailored for general terminal equipment. Given the constraints posed by limited hardware resources and diverse deployment environments, implementing such models constitutes a significant challenge. Therefore, the pursuit of more streamlined and stable models for carton appearance trademark detection emerges as a promising direction for future research.

In the context of real-world application, the scalability and adaptability of the proposed PTD-YOLO algorithm hold significant implications. While the results obtained are encouraging, further exploration is essential to advance the efficacy of these models in diverse operational environments, potentially incorporating adaptive mechanisms to optimize performance based on specific use-cases. The model's capabilities in accommodating hardware limitations and environmental variations set the stage for significant advancements in this field, lending potential for future explorations to take this work further. Lastly, the algorithm's capacity to effectively enhance detection accuracy in small target trademarks presents an important direction for subsequent studies, promising to drive more effective and nuanced understandings in the area of carton appearance trademark detection.

### REFERENCES

- [1] Biondi, F., Enescu, M.A., Given-Wilson, T., Legay, A., Nouredine, L., Verma, V. (2019). Effective, efficient, and robust packing detection and classification. *Computers & Security*, 85: 436-451. <https://doi.org/10.1016/j.cose.2019.05.007>
- [2] Nasrabadi, N.M. (2013). Hyperspectral target detection: An overview of current and future challenges. *IEEE Signal Processing Magazine*, 31(1): 34-44. <https://doi.org/10.1109/MSP.2013.2278992>
- [3] Angelova, A., Zhu, S. (2013). Efficient object detection and segmentation for fine-grained recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Portland, OR, USA, pp. 811-818. <https://doi.org/10.1109/CVPR.2013.110>
- [4] Gould, S., Gao, T., Koller, D. (2009). Region-based segmentation and object detection. *Advances in Neural Information Processing Systems*, 22: 655-663. <https://doi.org/10.5555/2984093.2984167>
- [5] Carreira, J., Sminchisescu, C. (2011). CPMC: Automatic object segmentation using constrained parametric min-cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(7): 1312-1328. <https://doi.org/10.1109/TPAMI.2011.231>
- [6] Paliwal, J., Visen, N.S., Jayas, D.S., White, N.D.G. (2003). Cereal grain and dockage identification using machine vision. *Biosystems Engineering*, 85(1): 51-57. [https://doi.org/10.1016/S1537-5110\(03\)00034-5](https://doi.org/10.1016/S1537-5110(03)00034-5)
- [7] Meng, Q., Qiu, R., He, J., Zhang, M., Ma, X., Liu, G. (2015). Development of agricultural implement system based on machine vision and fuzzy control. *Computers and Electronics in Agriculture*, 112: 128-138. <https://doi.org/10.1016/j.compag.2014.11.006>
- [8] Qureshi, W.S., Payne, A., Walsh, K.B., Linker, R., Cohen, O., Dailey, M.N. (2017). Machine vision for



- counting fruit on mango tree canopies. *Precision Agriculture*, 18: 224-244. <https://doi.org/10.1007/s11119-016-9458-5>
- [9] Gu, X., Yu, L., Pang, N., Martinez-Fernandez, J.S., Fu, X., Chen, S. (2020). Comparative techno-economic analysis of algal biofuel production via hydrothermal liquefaction: One stage versus two stages. *Applied Energy*, 259: 114115. <https://doi.org/10.1016/j.apenergy.2019.114115>
- [10] Girshick, R. (2015). Fast R-CNN. In *Proceedings of the IEEE International Conference on Computer Vision*, Santiago, Chile, pp. 1440-1448. <https://doi.org/10.1109/ICCV.2015.169>
- [11] He, K., Gkioxari, G., Dollár, P., Girshick, R. (2017). Mask R-CNN. In *Proceedings of the IEEE International Conference on Computer Vision*, Venice, Italy, pp. 2961-2969. <https://doi.org/10.1109/ICCV.2017.322>
- [12] Ren, S., He, K., Girshick, R., Sun, J. (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. *Advances in Neural Information Processing Systems*, 1: 91-99.
- [13] Jiang, P., Ergu, D., Liu, F., Cai, Y., Ma, B. (2022). A Review of Yolo algorithm developments. *Procedia Computer Science*, 199: 1066-1073. <https://doi.org/10.1016/j.procs.2022.01.135>
- [14] Guo, Y., Chen, S., Zhan, R., Wang, W., Zhang, J. (2022). LMSD-YOLO: A lightweight YOLO algorithm for multi-scale SAR ship detection. *Remote Sensing*, 14(19): 4801. <https://doi.org/10.3390/rs14194801>
- [15] Han, Z., Huang, H., Fan, Q., Li, Y., Li, Y., Chen, X. (2022). SMD-YOLO: An efficient and lightweight detection method for mask wearing status during the COVID-19 pandemic. *Computer Methods and Programs in Biomedicine*, 221: 106888. <https://doi.org/10.1016/j.cmpb.2022.106888>
- [16] Zhang, R., Wen, C. (2022). SOD-YOLO: A small target defect detection algorithm for wind turbine blades based on improved YOLOv5. *Advanced Theory and Simulations*, 5(7): 2100631. <https://doi.org/10.1002/adts.202100631>
- [17] Qi, C., Gao, J., Pearson, S., Harman, H., Chen, K., Shu, L. (2022). Tea chrysanthemum detection under unstructured environments using the TC-YOLO model. *Expert Systems with Applications*, 193: 116473. <https://doi.org/10.1016/j.eswa.2021.116473>
- [18] Siriani, A.L.R., Kodaira, V., Mehdizadeh, S.A., de Alencar Nääs, I., de Moura, D.J., Pereira, D.F. (2022). Detection and tracking of chickens in low-light images using YOLO network and Kalman filter. *Neural Computing and Applications*, 34(24): 21987-21997. <https://doi.org/10.1007/s00521-022-07664-w>
- [19] Niu, Z., Zhong, G., Yu, H. (2021). A review on the attention mechanism of deep learning. *Neurocomputing*, 452: 48-62. <https://doi.org/10.1016/j.neucom.2021.03.091>
- [20] Zhao, Z., Zhang, Q., Yu, X., Sun, C., Wang, S., Yan, R., Chen, X. (2021). Applications of unsupervised deep transfer learning to intelligent fault diagnosis: A survey and comparative study. *IEEE Transactions on Instrumentation and Measurement*, 70: 1-28. <https://doi.org/10.1109/TIM.2021.3116309>