Electronic Engineering and Informatics
G. Izat Rashed (Ed.)

doi:10.3233/ATDE240131

Capsule Defect Detection Method Based on an Improved Model

Jiasiyu ZHAO[1], Xindi DAI, Yu CHEN, Anqi ZHAO
School of Software Engineering, Jilin University, Changchun, 130012, China

Abstract. This article presents a novel method based on an enhanced version of the YOLOv5 model for detecting surface defects on capsules. The paper addresses the challenge of detecting defects on transparent capsules by introducing a deep learning-based approach called M-YOLO. Firstly, the backbone layer is replaced with MobileNetV3, enhancing the model's suitability for scenarios with limited storage space and power consumption. Secondly, a Cross-channel-H-SPP (CH-SPP) module is devised to augment the contextual information within the sensory field. To enhance defect detection accuracy, the SE attention mechanism is incorporated. Additionally, an improved label assignment strategy is employed to enhance the recall rate. Experimental results on the dataset demonstrate significant improvements in both accuracy and speed compared to the YOLOv5 model. The algorithm proposed in this article satisfies the requirement of processing every second (specific data).

Keywords. Image recognition, YOLOv5 network model, deep learning

1. Introduction

Capsules are extensively manufactured in the pharmaceutical industry. However, during the production process, the occurrence of defects in capsule products, such as dents, stains, and poor printing, poses significant challenges. Traditional defect detection methods in pharmaceutical plants rely on manual inspection or simple weighing, with manual inspection being susceptible to errors influenced by worker mood and physical fatigue [1].

Existing machine vision defect detection approaches commonly employ image processing techniques, including image enhancement and segmentation, to extract defect features from images. However, these methods often heavily rely on manually designed features and parameters [2]. With the advancement of deep learning techniques, neural networks have become a prominent tool for capsule surface defect detection [3]. Junlin Zhou et al. [4] proposed an improved Convolutional Neural Network (CNN) method called RACNN, which achieved high accuracy on a capsule dataset but exhibited low accuracy in recognizing deformed capsules. Zhiyuan Wang et al. [5] proposed an SVM-based complex component detection method that outperformed traditional CNN models in terms of accuracy. However, the model contained numerous parameters and incurred high computational complexity. To address these challenges, lightweight CNN models like MobileNetV2 [1] and MobileNetV3 [6] have emerged, featuring fewer parameters

[1] Corresponding author: Jiasiyu ZHAO, School of Software Engineering, Jilin University, e-mail: zjsylnfx@163.com

and lower computational requirements while maintaining high accuracy.

In recent years, target detection techniques [7] have gained popularity in the field of capsule surface defect detection, among which the YOLO series algorithm is widely adopted. However, the YOLO series algorithm still faces issues in capsule surface defect detection, including excessive computational resource demands, large model sizes, and false or missed detections [8]. To overcome these limitations, we propose an improved model based on YOLOv5. The contributions of this paper are as follows:

(1) We employ MobileNetV3, an efficient convolutional neural network model, to replace the backbone layer of YOLOv5, thereby enhancing detection accuracy. This modification renders the model more suitable for scenarios with limited storage space and power consumption.

(2) We introduce the Cross-channel-H-SPP module, which enhances spatial resolution and enriches contextual information within the sensory field.

(3) We incorporate an attention mechanism to enable the model to finely detect crucial parts.

In summary, this study addresses the limitations of existing methods by proposing improvements to the YOLOv5 model. These enhancements encompass the use of MobileNetV3, the Cross-channel-H-SPP module, an improved label assignment strategy, the SIOU loss function, and an attention mechanism, collectively leading to improved accuracy and performance in capsule surface defect detection.

2. Methodologies

This section mainly introduces the structure of Yolo, MobileNetv3, Cross-channel-H-SPP Hand the construction of M-YOLO.

2.1 YOLOv5

The YOLO (You Only Look Once) series models are widely employed for single-stage target detection. Its network architecture comprises four components: Input, Backbone, Neck, and Head. Figure 3 illustrates the detailed structure of the YOLOv5 module.

The Input module preprocesses the input images by resizing them to a size of 608×608. As shown in Figure 1, this preprocessing step scales the input image to match the network's input size and performs normalization operations. In addition, YOLOv5 introduces adaptive anchor box calculation and adaptive image scaling methods [9]. The adaptive anchor box calculation [10] sets a fixed anchor box size for the dataset. This enhancement method enriches the dataset, significantly improves the training speed of the network and reduces the memory requirements of the model [11]. Adaptive image scaling is used in the model inference process to avoid information redundancy and speed up the inference process.

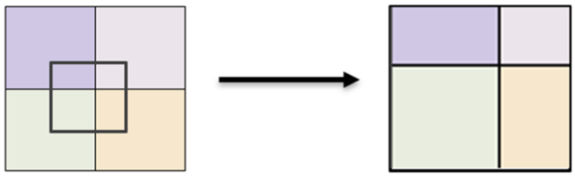

Figure 1. The Mosaic data enhancement process

The Backbone network consists of the Focus module and the CSP module, responsible for feature extraction from the input image. As described in Figure 2, the Focus module slices the input image data, reducing the height and width by half and increasing the number of channels to four times the original number. This transformation converts spatial information into channel information and reduces the number of floating-point operations. The CSP module divides the feature map of the base layer into two parts and merges them through a cross-stage hierarchy, ensuring accuracy while reducing computational requirements.

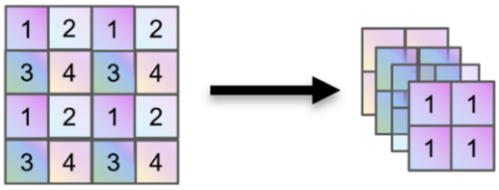

Figure 2. Focus slice operation

The Neck network fuses features from multiple scales to enhance the network's ability to detect objects of varying sizes [12]. While the Neck structure of YOLOv4 uses standard convolution operations for feature fusion [13], the Neck network of YOLOv5 incorporates the CSP2 structure from the CSPnet design, enhancing the network's feature fusion capabilities.

The Head network is responsible for predicting object classes and their associated bounding boxes. In YOLOv5, the Head network comprises multiple convolutional layers and prediction heads, which output the final detection results.

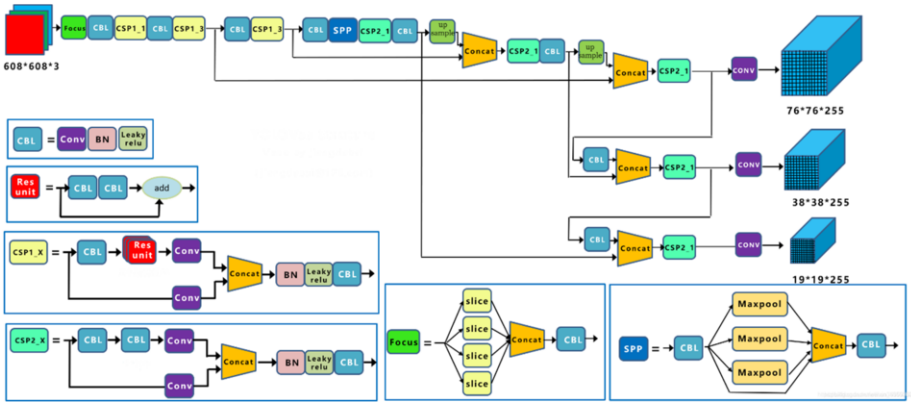

Figure 3. Structure of YOLOv5 network

2.2 MobileNetV3

MobileNetV3 is a lightweight network proposed by Google, primarily designed for mobile devices [6]. It introduces the lightweight activation function h-swish(x) [14]. The network structure is obtained through a combination of platform-aware NAS and NetAdapt [15] techniques. MobileNetV3 achieves excellent speed and accuracy while featuring fewer parameters, lower computation, and shorter inference time. These characteristics makeit suitable for scenarios with limited storage space and power

consumption, such as edge computing devices like mobile embedded devices. The core of the MobileNet model is the Depthwise separable convolution [16], which splits a normal convolution into a depthwise convolution and a pointwise convolution. The detailed structure is depicted in Figure 4. This approach allows for comparable results to standard convolution but with fewer parameters and operations.

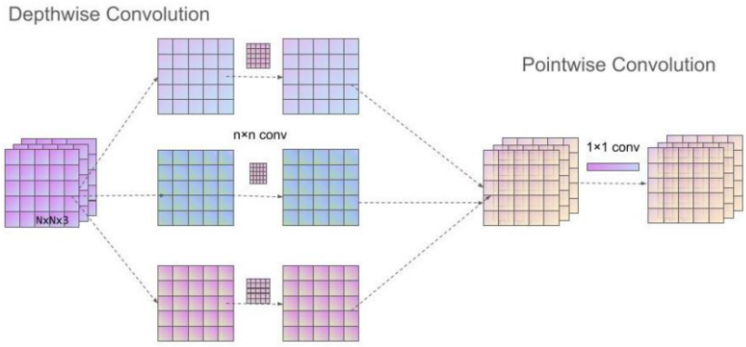

Figure 4. The deep separable convolution

2.3 Cross-channel-H-SPP

To capture contextual information and improve detection performance by understanding pixel relationships, it is crucial to incorporate contextual information. Inspired by the average-pooling method, Xu et al. [1] introduced a hybrid spatial pyramidal pooling module (H-SPP) [17]. Building upon H-SPP, we propose Cross-Channel-H-SPP (CH-SPP) and integrate it into the backbone layer. We perform average-pooling on the channel dimension to generate a new feature map, which we concatenate with the result of max-pooling.

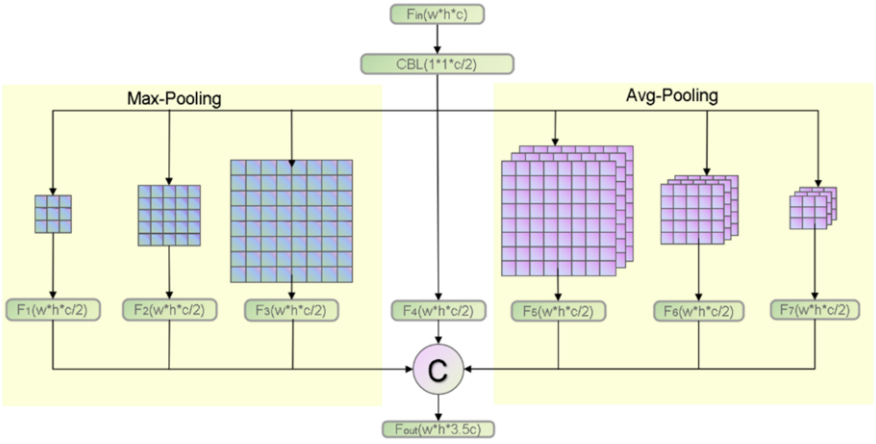

Figure 5. The detailed structure of CH-SPP module.

Figure 5 illustrates the detailed structure of the CH-SPP module, where "C" inside the circle denotes the concatenation operation. The CH-SPP module concatenates feature maps generated by two pooling layers. In the AveragePooling layer, we incorporate

neighboring images from the sample set to the pooling operation using kernels of different sizes, considering the correlation between preceding and subsequent images.

Firstly, the feature image "Fin" generated by the backbone layer passes through the CBL layer to generate an adapted channel image. Then, it is fed into both the MaxPooling layer and the AveragePooling layer with different kernel sizes (e.g., 3x3, 5x5, 8x8). Finally, the results are concatenated to form a new feature map .

2.4 M-YOLO

First, we replaced the backbone layer of YOLOv5 with MobileNetV3 to achieve a lighter network structure and improve both model accuracy and speed. MobileNetV3 is an efficient convolutional neural network that reduces computation and model size through lightweight network structures and depthwise separable convolutions. Then, we used the YOLOv5 detection head to detect surface defects on capsules.

The improved network structure can be divided into two main parts: the feature extraction network and the detection head. The feature extraction network adopts MobileNetV3, which possesses a lightweight network structure and employs depthwise separable convolutions. It comprises Inverted Residual Blocks (IRBs) and Linear Bottleneck Blocks (LBBs). IRBs extract low-level features from images, while LBBs extract high-level features. To adapt to the task of capsule surface defect detection, we added a Cross-channel-H-SPP module at the end of MobileNet Block to extract richer spatial and channel information. The detection head utilizes YOLOv5, which includes convolutional layers, pooling layers, and fully connected layers. The input to the detection head is the output of the feature extraction network, and the output comprises the detection results, including target bounding boxes and class predictions. We further improve detection accuracy by employing the SIOU loss function and attention mechanism. The improved algorithm's network architecture is depicted in Figure 6.

3. Experiment and Analysis

3.1 Experimental Data

We employed industrial cameras with pixel resolutions of 2448×1704, 1700×1100, 944×1024, and 944×944 to capture capsule images illuminated by large-diameter light sources. Due to the stringent yield rate requirements in actual pharmaceutical production, obtaining a sufficient amount of real defect samples was challenging. To mitigate overfitting and enhance the model's generalization ability, we performed data augmentation techniques on the dataset:

(1) Gamma transformation was applied to adjust the brightness of the images and improve the model's robustness under different light intensities.

(2) Contrast adjustment was performed using factor values of 0.8 and 1.2 to enhance the model's color registration robustness.

(3) Gaussian noise with a mean of 0 and variance of 0.02 was added to introduce image blurring and improve the model's sharpness robustness.

(4) We applied first-order gradient processing to highlight defects using the Sobel operator [18]. By processing the RGB three-channel images of each sample with the Sobel operator, we synthesized underlying feature maps to create a multi-channel input image.

A total of 1802 images were collected, with the training set comprising two-thirds of the data and the remaining one-third used for testing. The sample dataset included 95 capsule chips, 166 capsule depressions, and 421 capsule printing stains.

Following data augmentation, the number of images increased to 10,994.

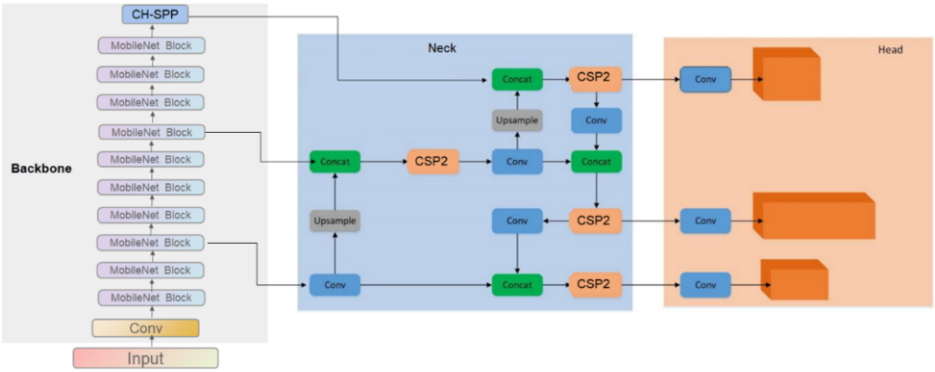

Figure 6. The M-YOLO network structure diagram

3.2 Experimental Environment and Parameters

In this paper, the experimental configuration and parameters are shown in Table 1.

Table 1. Experimental parameters table

Parameter	Value
batch-size	4
Image-size	608
lr	0.001
mosaic	0.8
mixup	0.243
momentum	0.843
GPU	NVIDIA GTX 1080 TI

3.3 Experimental Results

In this study, the model's detection accuracy was evaluated using mAP (mean Average Precision) and Recall, while the detection speed was measured using FPS (Frames Per Second). The Recall calculation formula is as follows:

$$Recall = \frac{TP}{TP + FN}$$

mAP@0.5 curve for the experiment is depicted in Figure 7. Table 2 presents a comparison of the experimental results between our proposed method and YOLOv5s. mAP@0.5 represents the mAP value with a IoU threshold of 0.5 or higher. A higher mAP value indicates greater model accuracy.

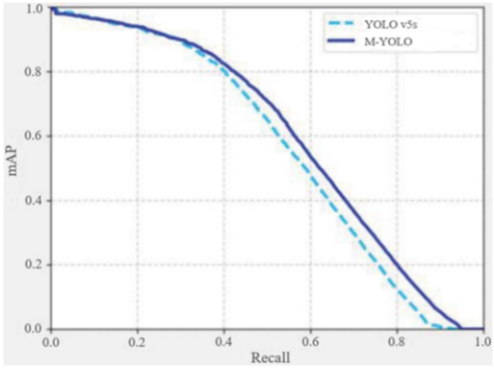

Figure 7. mAP@0.5-recall curves

Table 2. Experimental results of the comparison between M-YOLO and YOLOv5s

Method	mAP@0.5/%	FPS
YOLOv5s	76.75	19.25
Ours	83.79	18.60

Based on the above figure and table, it can be observed that the M-YOLO detection framework proposed in this paper achieves a detection speed of 18.60 FPS on the same dataset, slightly lower than YOLOv5s. However, in terms of detection accuracy, the M-YOLO algorithm outperforms YOLOv5s with a 7% higher mAP@0.5. In real industrial pharmaceutical scenarios, the proposed algorithm significantly enhances detection accuracy with an average detection time of 200ms per photo, making it highly suitable for capsule production scenarios.

Figure 8 illustrates the effect of partial defect detection. It can be seen that the proposed defect detection algorithm accurately detects various defects on the surface of capsules.

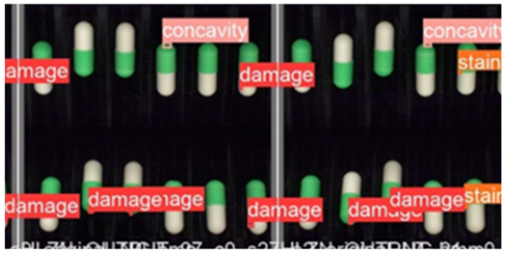

Figure 8. Examples of different types of defect detection results.

4. Conclusion

In conclusion, our proposed YOLOv5-MobileNetV3 approach has exhibited exceptional performance in the detection of capsule surface defects. It achieves an impressive mAP of 83.79% at an IoU threshold of 0.5, enabling accurate identification and localization of defects. The integration of MobileNetV3 and the Cross-channel-H-SPP module enhances the capture of spatial and channel information, facilitating effective detection of various defect types. Additionally, the incorporation of the SIOU loss function and attention

mechanism further enhances performance. These findings underscore the efficacy of our approach for quality control and manufacturing applications, highlighting its suitability for real-world defect detection tasks.

References

[1] Mark Sandler and Andrew Howard, "MobileNetV 2: Inverted Residuals and Linear Bottlenecks[C]", 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2018.
[2] Ren, Z., Fang, F., Yan, N. et al. State of the Art in Defect Detection Based on Machine Vision. Int. J. of Precis. Eng. and Manuf.-Green Tech. 9, 661–691 (2022).
[3] Chenghu Yuan. Research on image-based high-speed detection technology for internal defects of soft capsules [D]. Zhejiang Sci-Tech University,2019.DOI:10.27786/d.cnki.gzjlg.2019.000070.
[4] Junlin Zhou, Jiao He, Guoli Li, Yongbin Liu, "Identifying Capsule Defect Based on an Improved Convolutional Neural Network", Shock and Vibration, vol. 2020, Article ID 8887723, 9 pages, 2020. https://doi.org/10.1155/2020/8887723
[5] Wang Z, Zhu D. An accurate detection method for surface defects of complex components based on support vector machine and spreading algorithm[J]. Measurement, 2019, 147: 106886.
[6] Howard, Andrew, et al. "Searching for mobilenetv3." Proceedings of the IEEE/CVF international conference on computer vision. 2019.
[7] Janakiramaiah B, Kalyani G, Jayalakshmi A (2020) Automatic alert generation in a surveillance systems for smart city environment using deep learning algorithm. Evol, Intel
[8] Mohsin, Mazhar & Balogun, Oluwafemi & Haataja, Keijo & Toivanen, Pekka. (2023). Solid State Technology Volume: 66 Issue: 1 Publication Year. 66. 2023.
[9] Yao J, Qi J, Zhang J, Shao H, Yang J, Li X. A Real-Time Detection Algorithm for Kiwifruit Defects Based on YOLOv5. Electronics. 2021; 10(14):1711. https://doi.org/10.3390/electronics10141711
[10] Gao, M., Du, Y., Yang, Y. et al. Adaptive anchor box mechanism to improve the accuracy in the object detection system. Multimed Tools Appl 78, 27383–27402 (2019).
[11] Tai W, Wang Z, Li W, Cheng J, Hong X. DAAM-YOLOV5: A Helmet Detection Algorithm Combined with Dynamic Anchor Box and Attention Mechanism. Electronics. 2023; 12(9):2094.
[12] J. Zhou, W. Li, H. Fang, Y. Zhang and F. Pan, "The Hull Structure and Defect Detection Based on Improved YOLOv5 for Mobile Platform," 2022 41st Chinese Control Conference (CCC), Hefei, China, 2022, pp. 6392-6397, doi: 10.23919/CCC55666.2022.9902288.
[13] Souaidi, M., Ansari, M.E.: A new automated polyp detection network MP-FSSD in WCE and colonoscopy images based fusion single shot multibox detector and transfer learning. IEEE Access 10, 47124–47140 (2022)
[14] Z. You, H. Gao, S. Li, L. Guo, Y. Liu and J. Li, "Multiple Activation Functions and Data Augmentation-Based Lightweight Network for In Situ Tool Condition Monitoring," in IEEE Transactions on Industrial Electronics, vol. 69, no. 12, pp. 13656-13664, Dec. 2022, doi: 10.1109/TIE.2021.3139202.
[15] Junyi Chai, Hao Zeng, Anming Li, Eric W.T. Ngai,Deep learning in computer vision: A critical review of emerging techniques and application scenarios,Machine Learning with Applications,Volume 6,2021,100134,ISSN 2666-8270,https://doi.org/10.1016/j.mlwa.2021.100134.
[16] A G Howard, M Zhu, B Chen et al., MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications[J], 2017.
[17] S. Xu, H. Qian, W. Shen, F. Wang, X. Liu and Z. Xu, "Defect detection for PV Modules based on the improved YOLOv5s," 2022 China Automation Congress (CAC), Xiamen, China, 2022, pp. 1431-1436, doi: 10.1109/CAC57257.2022.10055183.
[18] Maini R, Aggarwal H. Study and comparison of various image edge detection techniques[J]. International journal of image processing (IJIP), 2009, 3(1): 1-11.