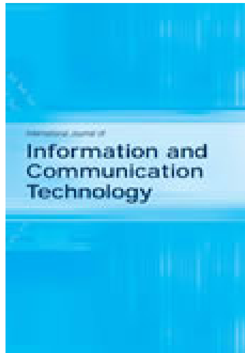
INDERSCIENCE PUBLISHERS
Linking academia, business and industry through research



International Journal of Information and Communication Technology

ISSN online: 1741-8070 - ISSN print: 1466-6642
https://www.inderscience.com/ijict

Wafer surface defect detection with enhanced YOLOv7

Chen Tang, Lijie Yin, Yongchao Xie

Article History:

Received:	28 March 2024
Last revised:	14 June 2024
Accepted:	02 July 2024
Published online:	12 September 2024

Copyright © 2024 Inderscience Enterprises Ltd.

Wafer surface defect detection with enhanced YOLOv7

Chen Tang, Lijie Yin* and Yongchao Xie

Hunan Railway Professional Technology College,
No. 18 Tianxin Avenue, Zhuzhou, Hunan Province, China
Email: 372384937@qq.com
Email: 271525290@qq.com
Email: xieyongchao2008@126.com
*Corresponding author

Abstract: Silicon wafers are crucial materials for semiconductor chip manufacturing. Detecting surface defects on wafers is essential for enhancing yield rates and identifying manufacturing issues. Traditional defect detection methods, relying on manual monitoring, are inefficient and inaccurate. Thus, there is a growing interest in leveraging deep learning for defect detection. However, existing algorithms still suffer from missed detections and slow processing speeds. To address these challenges, our study proposes a refined algorithm based on YOLOv7 for detecting wafer defects. We integrate SPD-Conv into the YOLOv7 MP module to enhance feature extraction accuracy and reduce computational complexity. Additionally, we incorporate the CBAM attention mechanism module into the backbone network to adapt to complex scenes. Moreover, we employ the SIoU loss function to improve bounding box regression accuracy. The WM-811k dataset is utilised for testing and evaluating the enhanced algorithm, achieving a recognition accuracy of 92.23%, a recall rate of 94.1%, and a mAP of 92.5%. Additionally, the frame rate remains stable at 136 frames per second, outperforming existing algorithms.

Keywords: surface defects on wafers; improved YOLOv7; SPD-Conv; CBAM attention mechanism; SIoU.

Reference to this paper should be made as follows: Tang, C., Yin, L. and Xie, Y. (2024) 'Wafer surface defect detection with enhanced YOLOv7', *Int. J. Information and Communication Technology*, Vol. 25, No. 6, pp.1–17.

Biographical notes: Chen Tang received her BSc degrees in Automation from South-Central University for Nationalities, China and MSc degree in Control Science from Hunan University of Science and Technology, China. Currently, he is a Lecturer at Hunan Railway Professional Technical College. His research interests include machine vision and deep learning.

Lijie Yin received his Bachelor's degree in Electronic Information Engineering from Hunan University of Technology and his Master's degree in Electronics and Communication Engineering from Xiangtan University. Currently, he is a Lecturer at Hunan Railway Professional Technology College, specialising in intelligent control.

Copyright © The Authors(s) 2024. Published by Inderscience Publishers Ltd. This is an Open Access Article distributed under the CC BY license. (http://creativecommons.org/licenses/by/4.0/)

Yongchao Xie received his BSc degrees in Communication Engineering from Lanzhou Jiaotong University, China and MSc degree in Communication and Information System from Southwest Jiaotong University, China. Currently, he is a Professor at Hunan Railway Professional Technical College, and he is a key young teacher in Hunan Province, China. His research field of centres on motor control technology.

1 Introduction

The semiconductor manufacturing sector has emerged as a critical industry, significantly impacting the development of multiple sectors. Notably, the production process of semiconductor wafers assumes a critical role in this context. The silicon wafer, serving as the initial material for semiconductor device fabrication, can accommodate thousands of minute grains in a single wafer. Each grain represents a chip to be processed, and the processing of silicon wafers is the foundational process in semiconductor chip manufacturing. However, during the transformation of silicon wafers into chips, intricate processes such as photolithography, polishing, etching, and cutting may engender various defects on the silicon wafer surface. If these defective grains are processed conventionally, the ultimate chip may manifest severe conductive defects, such as short circuits or open circuits, directly influencing the performance and quality of the chip. The main objective in detecting defects on wafers is to identify diverse types of defects, conduct a comprehensive analysis of their underlying causes, and expeditiously make adjustments to both equipment and operations, aiming to mitigate potential substantial losses (Kim et al., 2022).

The initial identification of surface defects on wafers is conventionally performed manually by seasoned inspectors. This undertaking requires a heightened level of concentration, and extended periods of visual scrutiny can result in visual fatigue and diminished attentiveness. The manual inspection process presents challenges including slow inspection speed, diminished accuracy, elevated costs, and subjective assessments (Ma et al., 2023). Researchers employ carefully designed imaging systems to acquire uniformly illuminated images, enabling inspection personnel to observe surface defects clearly (Cumbajin et al., 2023; Zhao et al., 2024; Kim and Behdinan, 2023; Theodosiou et al., 2023).

The advancement of image processing technology has led to the application of image processing-based algorithms in wafer surface defect detection (Jha and Babiceanu, 2023; Jizat et al., 2021; Xia et al., 2023; Yang et al., 2020). Post wafer imaging, template images are aligned with wafer images to eliminate background information. Subsequently, defect areas are categorised using manually designed features. Nevertheless, in real-world production scenarios, acquiring template images poses a challenge, and variations in brightness, angles, and proportions between wafer images and templates are common. This discrepancy can result in erroneous defect reports.

In recent years, the efficiency of deep learning methodologies, particularly convolutional neural networks (CNNs), has significantly advanced (Simonyan and Zisserman, 2014; He et al., 2016; Huang et al., 2017), across various computer vision realms, abundant surface defect detection (SDD) approaches leveraging CNNs have emerged (Wang et al., 2022). CNN-based networks simplify the algorithm development process by directly inputting wafer images into the model, eliminating the need for intricate image processing steps. This streamlined approach not only reduces the complexity of algorithm development but also enhances accuracy. For example, Krizhevsky et al. (2012) utilised AlexNet, Wang et al. (2020) introduced a hybrid DC-Net, termed as MDPR. Wang et al. (2020) conducted tests on six different wafer image databases using a combined classifier consisting of MPL, SVM, and CNN. Nakazawa and Kulkarni (2018) presented an approach involving the use of synthetically generated data to train and validate CNN models with simulated wafer images.

In certain instances, possessing knowledge solely about the categories is insufficient; there is a need to ascertain their locations for a more precise identification of fault causes. Consequently, numerous object detection algorithms are utilised in wafer surface defect detection. Object detection networks are primarily split into two categories, with the first category encompasses the two-stage networks. Two-stage networks are distinguished by their utilisation of a region proposal network to produce candidate boxes initially, which are subsequently subjected to classification and regression. A classic example of a two-stage network is faster R-CNN (Girshick, 2015). Zheng and Zhang (2023) introduced an innovative approach for detecting surface defects on semiconductor wafers, utilising background subtraction in conjunction with faster R-CNN. Furthermore, to bolster feature extraction capabilities, a spatial attention module was integrated into the network, thereby improving detection performance, particularly for larger defects. The second type is the one-stage network, dispensing with the need for candidate box generation and instead directly providing classification and regression information for target objects. Typically, one-stage networks exhibit faster detection speeds. Chen et al. (2020a) pioneered the integration of generative adversarial networks (GANs) with the YOLOv3 object detection algorithm for detecting wafer defects in limited samples, enhancing YOLOv3's capacity for generalisation. In detecting and locating wafer defects, Shinde et al. (2022) recommended the utilisation of the upgraded YOLOv4. In comparison to YOLOv3, YOLOv4 demonstrates significantly enhanced detection capabilities, delivering more efficient detection and localisation performance for intricate wafer defect patterns. Presently, the YOLO algorithm is extensively employed in industrial production and has emerged as the predominant object detection algorithm. Consequently, this paper adopts the most outstanding comprehensive performance algorithm in the YOLO series – YOLOv7, as proposed by Wang et al. (2023). Simultaneously, acknowledging the characteristics of small target feature sizes and the diverse and interwoven defect types on the wafer surface, targeted improvements are made to YOLOv7 in this study.

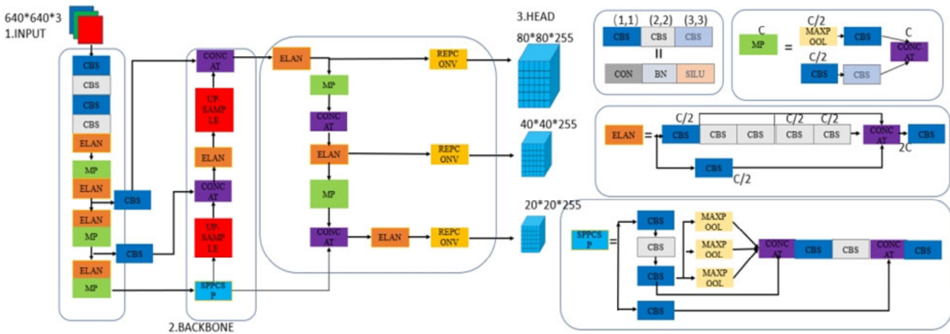
2 Related work

YOLOv7 is short for you only look once version 7, which is an improved version compared to YOLOv5. In comparison to its predecessor, YOLOv7 undergoes

optimisations in terms of network structure, data augmentation, and other facets, with the objective of achieving superior and efficient object detection. YOLOv7 exhibits precision in accurately detecting the positions and categories of multiple objects, leading to its widespread application in intelligent surveillance and industrial inspection (Chen and Dang, 2023; Chen et al., 2023; Qiu et al., 2023; Yu et al., 2023; Yuan et al., 2022). YOLOv7 consists of various components, Figure 1 illustrates the network structure visually.

- Input layer: the input layer resizes input graphics to 640 × 640, followed by random cropping, scaling, and concatenation to achieve data augmentation.
- Backbone network: feature extraction is the main role of the backbone, utilising a customised version of CSPNet as the backbone network, incorporating of CBS, ELAN, and MP architectures is integral. CBS, which comprises Conv+BN+SiLU, is primarily utilised for channel transformation, feature extraction, and image subsampling. ELAN, consisting of multiple CBS units, ensures consistency in input-output feature sizes. The channel numbers vary in the first two CBS units, while the subsequent ones maintain consistent input and output channels, ultimately outputting the required channels after the last CBS. The MP module, utilised for detecting objects across different scales successfully pinpoints items of diverse dimensions.
- Detection head: the detection head employs SPPCSP and ELAN modules to aggregate image features, retaining three detection heads for predicting target object class probabilities, confidence scores, and bounding box coordinates. The detection heads produce features at three scales: 20 × 20, 40 × 40, and 80 × 80, corresponding to large, medium, and small targets, respectively.

Figure 1 Structure of the YOLOv7 network (see online version for colours)



3 Proposed method

3.1 SPD-Conv

When detecting defects on the surface of semiconductor wafers, detecting tiny defects poses a challenge due to their often elusive nature. The difficulty arises from the low pixel count of these small target objects, resulting in limited information for the model to

learn. Convolutional neural networks effectively filter out a significant amount of redundant information within the receptive field. However, in cases where the image pixel count is low, the utilisation of strided convolutions and pooling could potentially result in the diminishing of detailed information. The space-to-depth convolution (SPD-Conv), with its convolutional operation and treatment of symmetric positive definite matrices, offers superior capabilities for extracting and utilising intrinsic features of the data. It demonstrates enhanced feature extraction while maintaining the count of parameters minimal and the computational intricacy at a reduced level, resulting in faster computations. Therefore, our approach integrates the spatial pyramid detection-convolutions (Jeyaprakash et al., 2020; Sunkara and Luo, 2022; Li et al., 2023) into the MP module of YOLOv7, thereby enhancing the precision of defect identification. The subsequent example will serve to clarify our point. Taking into account an input image X that has the dimensions of $W \times W \times D1$, we proceed to analyse how, at each stride interval, it is divided into multiple subsequent subsequences, each capturing distinct sub-features, as shown below:

$$f_{0,0} = X[0:W:P, 0:W:P], \tag{1}$$

$$f_{1,0} = X[1:W:P, 0:W:P], \ldots, \tag{2}$$

$$f_{P-1,0} = X[P-1:W:P, 0:W:P] \tag{3}$$

$$f_{0,1} = X[0:W:P, 1:W:P], \tag{4}$$

$$f_{1,1} = X[1:W:P, 1:W:P], \ldots, \tag{5}$$

$$f_{P-1,1} = X[P-1:W:P, 1:W:P] \tag{6}$$

$$f_{0,P-1} = X[0:W:P, P-1:W:P],\ f1, P-1, \ldots, \tag{7}$$

$$f_{P-1,P-1} = X[P-1:W:P, P-1:W:P] \tag{8}$$

where W is the size of the input image X and P is the partition factor, which is used to calculate the output of the SPD layer.

As an example, consider an arbitrary intermediate feature map X. Each sub-map $f(x, y)$ is constructed from feature vectors $X(x, y)$, designed such that the coordinates $x + i$ and $y + j$ are scaled in a proportional manner. Consequently, a down-sampling of the feature map X by a certain factor is achievable by each sub-map. An instance is depicted in Figure 2, where P is set to 2, demonstrating the formation of the ensuing sub-feature sequences as follows:

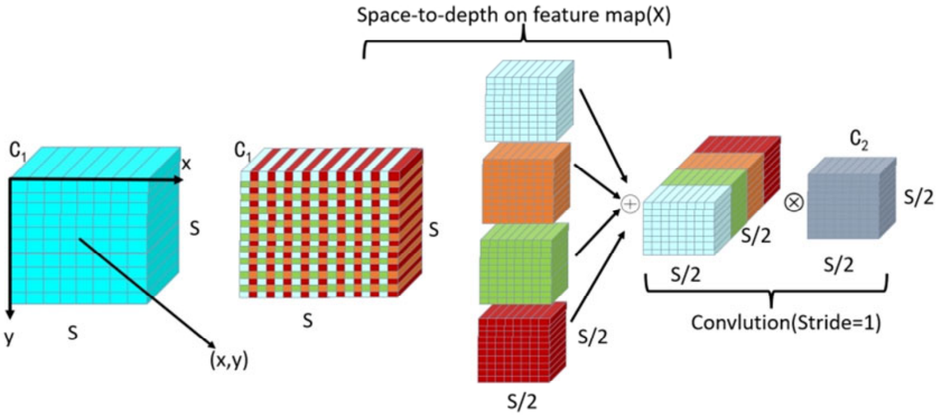
$$f_{0,0} = X[0:W:P, 0:W:P] \tag{9}$$

$$f_{1,0} = X[1:W:P, 0:W:P] \tag{10}$$

$$f_{0,1} = X[0:W:P, 1:W:P] \tag{11}$$

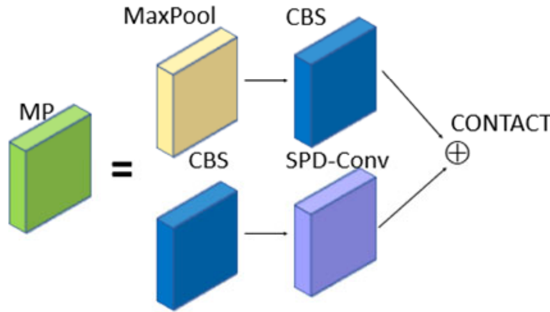
$$f_{1,1} = X[1:W:P, 1:W:P] \tag{12}$$

Figure 2 Illustration of SPD-Conv structure with partition factor $P = 2$ (see online version for colours)



Each sub-map takes the form of (W/2, W/2, D1), employing a downsampling factor set to 2. Ultimately, these sub-feature maps are merged along the channel axis, resulting in the feature map X', which exhibits a spatial reduction by a multiple of P and an increase in channel dimensions by a factor of P2. As the MP module contains convolution operations with a stride of 2, this study replaced the strided convolutions in the feature fusion network with SPD-Conv, as illustrated in Figure 3.

Figure 3 Improvement of MP module (see online version for colours)



3.2 *Add attention mechanism*

In surface defect detection on semiconductor wafers, the defects to be detected are often small. In order to augment the model's capacity for detecting such targets, the YOLOv7 backbone network's ultimate layer has been supplemented with the convolutional block attention module (CBAM) (Woo et al., 2018; Christou et al., 2004; Canayaz, 2021), as illustrated in Figure 4. CBAM comprises two distinct sub-modules – the channel attention module that compresses spatial dimensions while retaining channel information to focus on meaningful image features, and a spatial attention module that directs attention towards object location to adjust resource allocation based on feature importance (Chen et al., 2020b). By simultaneously allocating attention across channel

and spatial dimensions, CBAM improves model performance (Guo et al., 2023). The basic CBAM structure is illustrated in Figure 5.

Figure 4 YOLOv7 backbone network with added CBAM module (see online version for colours)

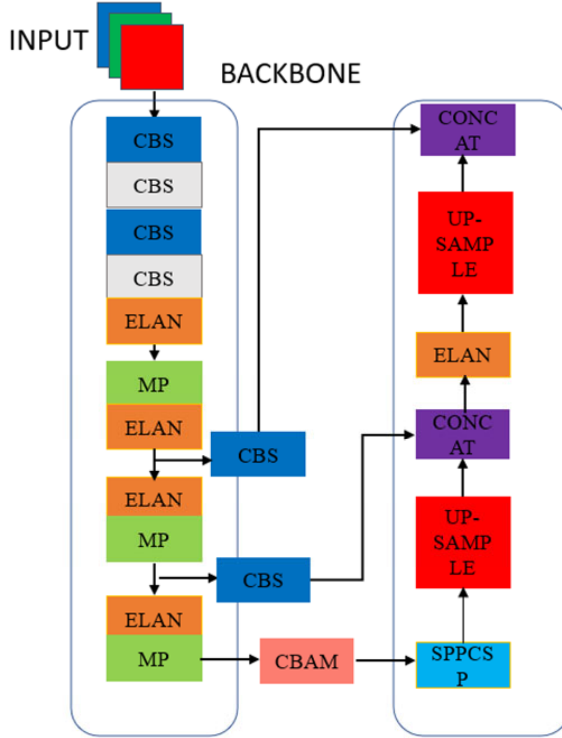
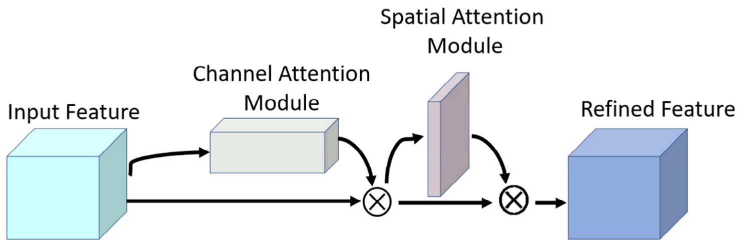


Figure 5 Structure of CBAM (see online version for colours)



The overall process of CBAM for feature maps generated by the network backbone includes the following steps:

$$F \in R^{C \times H \times W} \tag{13}$$

Attention maps of the 1D channel type are produced by CBAM through the following process:

$$M_C \in R^{C \times 1 \times 1} \tag{14}$$

Spatial attention feature map in two dimensions:

$$M_S \in R^{1\times H\times W} \tag{15}$$

This process can be delineated through the subsequent formula:

$$F' = M_C(F) \otimes F \tag{16}$$

$$F'' = M_S(F') \otimes F' \tag{17}$$

where C, H and W represent the number of channels, height, and width, respectively.

3.3 Optimisation loss function

YOLOv7's loss function comprises three distinct elements (Liu et al., 2023). The loss associated with bounding boxes evaluates the discrepancy in localisation between the predicted and the actual ground truth boxes. The target confidence loss quantifies the error in predicting whether the object in a box is a target. The class loss assesses the error in predicting object categories within boxes. YOLOv7 employs BCEWithLogitsLoss for the target confidence and class loss functions, while adopting the loss pertaining to bounding boxes utilises CIoU (Du et al., 2021) metric. The CIoU loss is defined in equation (18).

$$Loss_{CIoU} = 1 - IoU + \frac{\rho^2(b, b^{gt})}{c^2} + \alpha v \tag{18}$$

In this context, $\rho^2(b, b^{gt})$ represents the distance, measured in the Euclidean sense, between the centroids of the predicted bounding box and the corresponding ground truth bounding box. The symbol 'c' corresponds to the length of the diagonal of the tiniest enclosing rectangle containing both the ground truth and predicted boxes. The calculation for α is given by equation (12), and the calculation for v is given by equation (19).

$$\alpha = \frac{v}{1 - IoU + v} \tag{19}$$

$$v = \frac{4}{\pi^2}\left(\arctan\frac{w^{gt}}{h^{gt}} - \arctan\frac{w}{h}\right)^2 \tag{20}$$

The loss function based on CIoU does not completely consider the variances in orientation between the bounding boxes that are predicted and those that are grounded in truth. To address this, we introduce the SIoU (Gevorgyan, 2022) as a new bounding box similarity measurement that considers angles. In the equation, the CIoU loss has been substituted with the SIoU loss to improve the network's efficacy. The SIoU loss is composed of four distinct elements:

1 Angle loss

The angle loss calculation method is shown in equation (21).

$$\Lambda = \cos\left(2\times\left(\arcsin\left(\frac{c_h}{\sigma}\right) - \frac{\pi}{4}\right)\right) \tag{21}$$

$$c_h = \max\left(b_{cy}^{gt}, b_{cy}\right) - \min\left(b_{cy}^{gt}, b_{cy}\right) \quad (22)$$

$$\sigma = \sqrt{\left(b_{cx}^{gt}, b_{cx}\right)^2 - \left(b_{cy}^{gt}, b_{cy}\right)^2} \quad (23)$$

where $(b_{cx}^{gt}, b_{cy}^{gt})$ representing the central point of the actual bounding box and (b_{cx}, b_{cy}) indicating the central point of the inferred bounding box.

2 Distance loss

The calculation method for distance loss is shown in equation (24).

$$\Delta = 2 - e^{-\lambda\rho x} - e^{-\lambda\rho y} \quad (24)$$

$$\rho_x = \left(\frac{b_{c_x}^{gt} - b_{c_x}}{c_w}\right)^2, \quad \rho_y = \left(\frac{b_{c_y}^{gt} - b_{c_y}}{c_h}\right)^2 \quad (25)$$

$$\gamma = 2 - \Lambda \quad (26)$$

(c_w, c_h) denotes the dimensions of the smallest rectangle that encompasses both the projected and the actual bounding boxes.

3 Shape loss

The calculation method for shape loss is presented within the confines of equation (27).

$$\Omega = \left(1 - e^{-w_w}\right)^\theta + \left(1 - e^{-w_h}\right)^\theta \quad (27)$$

$$w_w = \frac{\left|w - w^{gt}\right|}{\max\left(w, w^{gt}\right)}, \quad w_h = \frac{\left|h - h^{gt}\right|}{\max\left(h, h^{gt}\right)} \quad (28)$$

4 IoU loss

The IoU loss computation is delineated in equation (29).

$$IoU = \frac{A}{B} \quad (29)$$

A denotes the area of intersection between the actual bounding box and the one that has been projected, while B denotes the area of union between the two boxes. When calculating IoU, the ratio A/B signifies the degree of match between the two boxes.

5 SIoU loss

The method for computing the SIoU loss function used in this algorithm is defined in equation (30).

$$Loss_{SIoU} = 1 - IoU + \frac{\Delta + \Omega}{2} \quad (30)$$

4 Experiment and analysis

4.1 Experimental setting

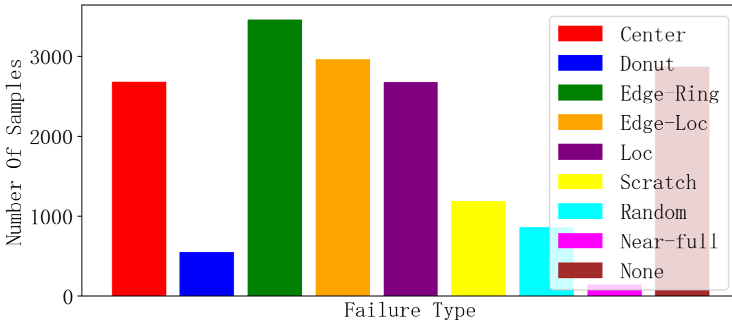
The model proposed within the paper was implemented and tested using Python 3.8 and PyTorch 1.11.0. The computational setup used for experimentation consists of: Ubuntu 16.04 LTS operating system, Intel Xeon(R) E5-2680 v4 processor (3.3 GHz, 14 cores), and Nvidia Tesla T4 GPU (16 GB memory). Table 1 provides details regarding the specific parameters of the experimental environment.

Table 1 Parameters of experimental environment

Component	*Name*
Operating system	Ubuntu 16.04 LTS
CPU	Intel Xeon(R) E5-2680 v4
GPU	Nvidia Tesla T4
Training acceleration	CUDA 11.7
Programming language	Python 3.8
Deep learning framework	PyTorch 1.11.0

In this study, we utilised the publicly available WM-811k dataset (Yu et al., 2019) for both training and testing. This dataset comprises 811,457 wafer images collected from an actual semiconductor manufacturing process, with only 20% labelled. Due to the imbalanced distribution of defects, we uniformly selected 8 of the defect classes along with the defect-free images, totalling 18,143 wafer images for model development, as illustrated in Figure 6. Specifically, a total of 80% was allocated for the training phase, with 10% set aside for validation and the remaining 10% reserved for testing purposes. The nine types of defects in the dataset are illustrated in Figure 7.

Figure 6 Number of samples for different types of defects (see online version for colours)



4.2 Evaluation parameters

To assess algorithm performance, we employed multiple evaluation metrics validating different aspects. Frames per second (FPS) reflects real-time detection capability. Precision refers to the ratio of true positives to positive predictions, while recall denotes the proportion of correctly identified actual positives. Equations (31) and (32) define

precision and recall, with TP representing true positives, FP representing false positives, and FN representing false negatives. AP represents per-class accuracy, averaged across recall thresholds to compute mean average precision (mAP) in equation (34). This evaluates accuracy across all classes, defined in equation (33).

$$Precision = \frac{TP}{TP + FP} \tag{31}$$

$$Recall = \frac{TP}{TP + FN} \tag{32}$$

$$AP = \int_0^1 P(r)dr \tag{33}$$

$$mAP = \frac{1}{N}\sum_{1}^{N} AP_i \tag{34}$$

4.3 Experimental analysis

For the purpose of evaluating the detection efficacy of the enhanced YOLOv7 network architecture in comparison to current models employed in object detection, we carry out a comparative analysis with the currently other classical object detection models. The evaluation metrics for each model are summarised in Table 2. Observing the table allows us to note that the mAP of improved-YOLOv7 reaches 92.5%, significantly higher than other methods, while also exhibiting superior recall and precision values. Furthermore, we utilise FPS as an assessment metric for object detection speed. From the FPS metric, it is evident that our method achieves high detection speed, particularly outperforming faster R-CNN by a factor of 6. Overall, our method demonstrates high accuracy for semiconductor wafer surface defect detection, coupled with high detection efficiency.

In Figure 8, a comparative evaluation is presented, depicting the overall loss curves of the improved-YOLOv7 model alongside those of other traditional models for object detection. The improved-YOLOv7 loss curve exhibits a faster decreasing trend, stable convergence speed, and lower final loss value compared to the other models. This indicates that the improved-YOLOv7 model exhibits a higher degree of training proficiency and operational excellence. Figures 9, 10, and 11 depict a comparative analysis of the training procedures for the improved-YOLOv7 model with other traditional models for object detection. This analysis indicates that during the initial training phase, improved-YOLOv7 demonstrates robust learning capability and achieves faster convergence. Additionally, it exhibits better mAP, recall, and precision values. The comparisons above underscore the strong performance of improved-YOLOv7.

In Figures 12 and 13, the detection performance of the Improved YOLOv7 is compared with that of the original YOLOv7. The graphs demonstrate the improved YOLOv7 model achieves higher confidence in detecting wafer defects. Additionally, when the confidence threshold was set to 0.1 and the NMS IoU threshold was set to 0.1, we observed that the original YOLOv7 model missed detections of small defects, whereas the improved YOLOv7 successfully detected these small defects. This further confirms the suggested method’s effectiveness at lowering the incidence of undetected cases.

Table 2 Performance comparison with other classical object detection models

Method	*Precision*	*Recall*	*mAP*	*FPS*
Improved YOLOv7	92.23%	94.1%	92.5%	136
YOLOv7	88.56%	87.4%	86.2%	116
YOLOv5	83.36%	86.8%	84.1%	91
Faster-RCNN	89.06%	92.1%	89.2%	22

Figure 7 Example semiconductor wafer defects in WM-811k dataset, (a) center (b) donut (c) EdgeLoc (d) EdgeRing (e) loc (f) random (g) scratch (h) near-full (i) none (see online version for colours)

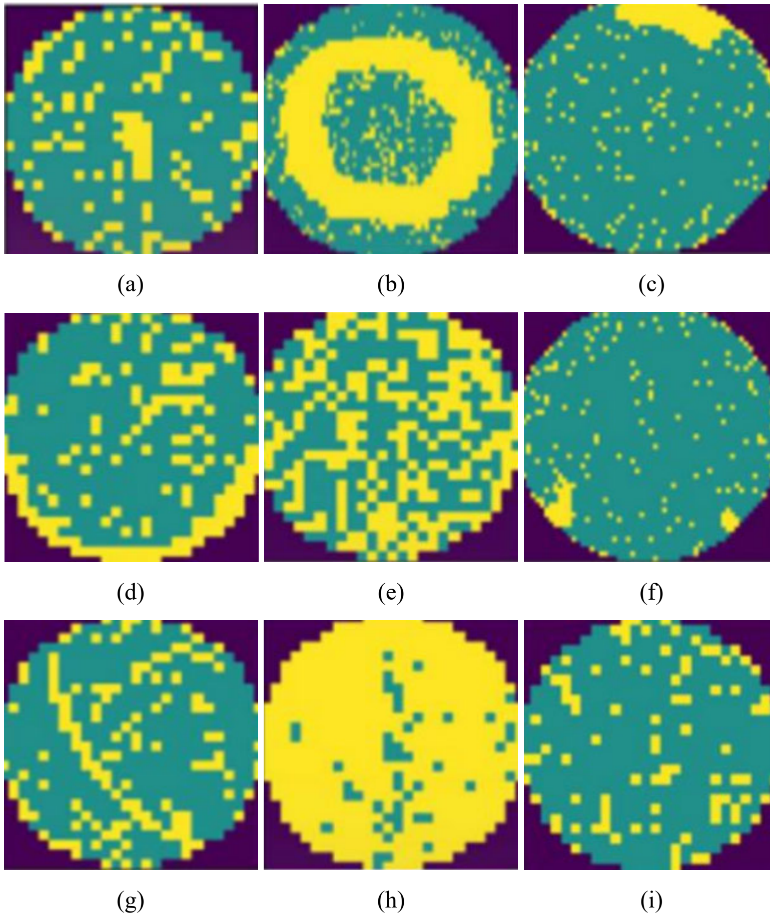


Figure 8 Comparative analysis of total loss curves between improved-YOLOv7 and traditional object detection models (see online version for colours)

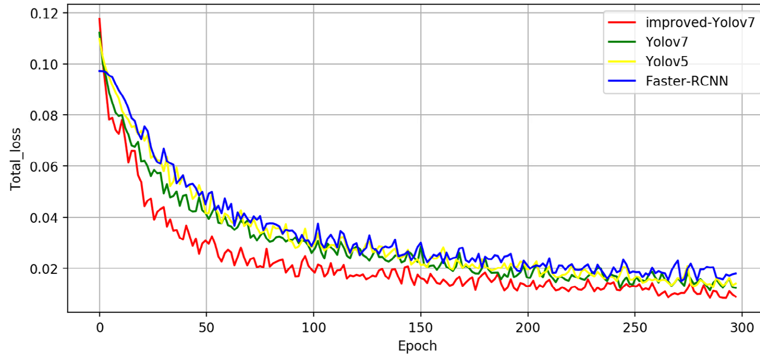


Figure 9 Comparison of recall curves for improved-YOLOv7 and classical object detection models (see online version for colours)

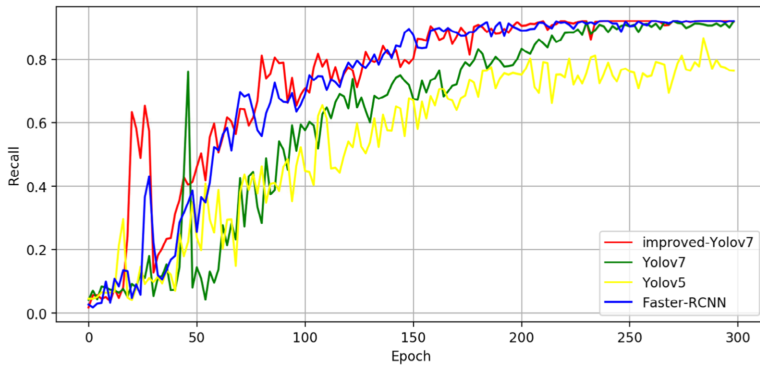


Figure 10 Comparison of precision curves for improved-YOLOv7 and classical object detection models (see online version for colours)

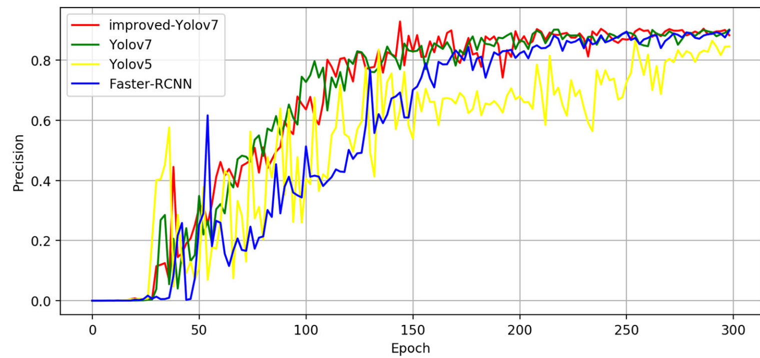


Figure 11 Comparison of MAP curves for improved-YOLOv7 and classical object detection models (see online version for colours)

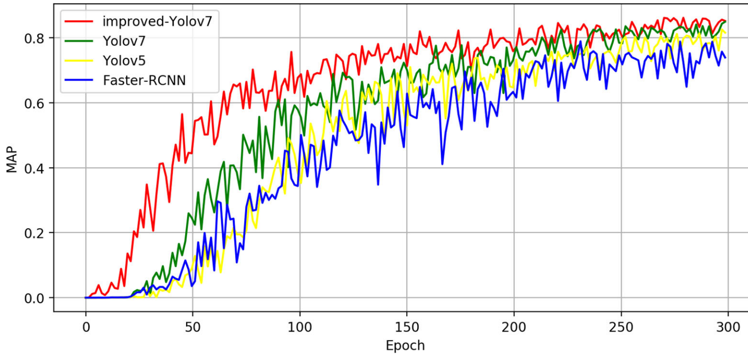


Figure 12 Detection results of YOLOv7 (see online version for colours)

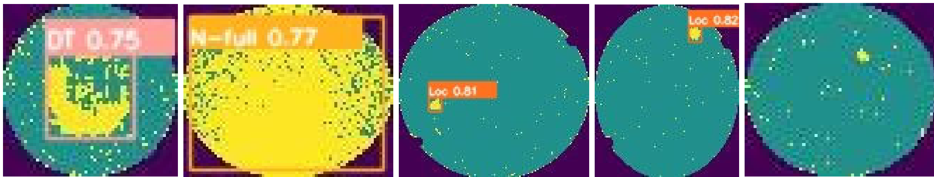


Figure 13 Detection results of improved-YOLOv7 (see online version for colours)

5 Conclusions

This study aims to address challenges in defect localisation and classification in semiconductor manufacturing processes by the improved YOLOv7 algorithm. The YOLOv7 MP module is enhanced through the integration of SPD-Conv in the improved algorithm to improve feature extraction accuracy and reduce computational complexity. Additionally, the YOLOv7 network is enhanced by embedding the CBAM attention mechanism module, thereby boosting its capacity for feature extraction and enhancing detection capabilities for smaller and less prominent targets. Furthermore, to accelerate training speed and enhance inference accuracy, we substitute the bounding box loss function in YOLOv7 with SIoU.

The experimental outcomes indicate that the suggested algorithm attains a high level of precision on the WM-811k dataset, with a mAP of 92.5%, while also enhancing detection speed to 136 frames per second. As semiconductor chip features become increasingly smaller, traditional detection methods struggle to effectively identify wafer

surface defects. However, the results still indicate a significant proportion of misjudgements and omissions, highlighting the need for further optimisation of network architectures in future work.

Acknowledgements

The research received support from the Scientific Research Projects of the Hunan Provincial Department of Education (No. 23C0806) and Scientific Research and Innovation Team Construction Project of Hunan Railway Professional Technology College.

References

Canayaz, M. (2021) 'C+EffxNet: a novel hybrid approach for COVID-19 diagnosis on CT images based on CBAM and EfficientNet', *Chaos, Solitons & Fractals*, Vol. 151, p.111310.

Chen, B. and Dang, Z. (2023) 'Fast PCB defect detection method based on FasterNet backbone network and CBAM attention mechanism integrated with feature fusion module in improved YOLOv7', *IEEE Access*, Vol. 11, pp.95092–95103.

Chen, J., Ma, B., Ji, C. et al. (2023) 'Apple inflorescence recognition of phenology stage in complex background based on improved YOLOv7', *Computers and Electronics in Agriculture*, Vol. 211, p.108048.

Chen, S.H., Kang, C.H. and Perng, D.B. (2020a) 'Detecting and measuring defects in wafer die using GAN and YOLOv3', *Appl. Sci.*, Vol. 10, No. 23, p.8725.

Chen, Y., Zhang, X., Chen, W. et al. (2020b) 'Research on recognition of fly species based on improved RetinaNet and CBAM', *IEEE Access*, Vol. 8, pp.102907–102919.

Christou, C., Eliophotou-Menon, M. and Philippou, G. (2004) 'Teachers' concerns regarding the adoption of a new mathematics curriculum: an application of CBAM', *Educational Studies in Mathematics*, Vol. 57, pp.157–176.

Cumbajin, E., Rodrigues, N., Costa, P. et al. (2023) 'A systematic review on deep learning with CNNs applied to surface defect detection', *Journal of Imaging*, Vol. 9, No. 10, p.193.

Du, S., Zhang, B., Zhang, P. et al. (2021) 'An improved bounding box regression loss function based on CIOU loss for multi-scale object detection', *2021 IEEE 2nd International Conference on Pattern Recognition and Machine Learning (PRML)*, IEEE, pp.92–98.

Gevorgyan, Z. (2022) *SIoU Loss: More Powerful Learning for Bounding Box Regression*, arXiv preprint arXiv:2205.12740.

Girshick, R. (2015) 'Fast R-CNN', *Proceedings of the IEEE International Conference on Computer Vision*, pp.1440–1448.

Guo, Y., Aggrey, S.E., Yang, X. et al. (2023) 'Detecting broiler chickens on litter floor with the YOLOv5-CBAM deep learning model', *Artificial Intelligence in Agriculture*, Vol. 9, pp.36–45.

He, K., Zhang, X., Ren, S. and Sun, J. (2016) 'Deep residual learning for image recognition', *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp.770–778.

Huang, G., Liu, Z., van der Maaten, L. and Weinberger, K.Q. (2017) 'Densely connected convolutional networks', *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp.4700–4708.

Jeyaprakash, S., Heffernan, J.E., Driscoll, R.H. et al. (2020) 'Impact of drying technologies on tomato flavor composition and sensory quality', *LWT*, Vol. 120, p.108888.

Jha, S.B. and Babiceanu, R.F. (2023) 'Deep CNN-based visual defect detection: survey of current literature', *Computers in Industry*, Vol. 148, p.103911.

Jizat, J.A.M., Majeed, A.P.P.A., Nasir, A.F.A. et al. (2021) 'Evaluation of the machine learning classifier in wafer defects classification', *ICT Express*, Vol. 7, No. 4, pp.535–539.

Kim, T. and Behdinan, K. (2023) 'Advances in machine learning and deep learning applications towards wafer map defect recognition and classification: a review', *Journal of Intelligent Manufacturing*, Vol. 34, No. 8, pp.3215–3247.

Kim, T.S., Lee, J.W., Lee, W.K. et al. (2022) 'Novel method for detection of mixed-type defect patterns in wafer maps based on a single shot detector algorithm', *Journal of Intelligent Manufacturing*, Vol. 33, No. 6, pp.1715–1724.

Krizhevsky, A., Sutskever, I. and Hinton, G.E. (2012) Imagenet classification with deep convolutional neural networks', *Advances in Neural Information Processing Systems*, Vol. 60, No. 6, pp.84–90.

Kyeong, K. and Kim, H. (2018) 'Classification of mixed-type defect patterns in wafer bin maps using convolutional neural networks', *IEEE Transactions on Semiconductor Manufacturing*, Vol. 31, No. 3, pp.395–402.

Li, A., Sun, S., Zhang, Z. et al. (2023) 'A multi-scale traffic object detection algorithm for road scenes based on improved YOLOv5', *Electronics*, Vol. 12, No. 4, p.878.

Liu, Q., Wu, T., Deng, Y. et al. (2023) 'SE-YOLOv7 landslide detection algorithm based on attention mechanism and improved loss function', *Land*, Vol. 12, No. 8, p.1522.

Ma, J., Zhang, T., Yang, C. et al. (2023) 'Review of wafer surface defect detection methods', *Electronics*, Vol. 12, No. 8, p.1787.

Nakazawa, T. and Kulkarni, D.V. (2018) 'Wafer map defect pattern classification and image retrieval using convolutional neural network', *IEEE Transactions on Semiconductor Manufacturing*, Vol. 31, No. 2, pp.309–314.

Qiu, Y., Lu, Y., Wang, Y. et al. (2023) 'IDOD-YOLOV7: image-dehazing YOLOV7 for object detection in low-light foggy traffic environments', *Sensors*, Vol. 23, No. 3, p.1347.

Shinde, P.P., Pai, P.P. and Adiga, S.P. (2022) 'Wafer defect localization and classification using deep learning techniques', *IEEE Access*, Vol. 10, pp.39969–39974.

Simonyan, K. and Zisserman, A. (2014) *Very Deep Convolutional Networks for Large-Scale Image Recognition*, arXiv preprint arXiv:1409.1556.

Sunkara, R. and Luo, T. (2022) 'No more strided convolutions or pooling: a new CNN building block for low-resolution images and small objects', *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, Springer Nature Switzerland, Cham, pp.443–459.

Theodosiou, T., Rapti, A., Papageorgiou, K. et al. (2023) 'A review study on ML-based methods for defect-pattern recognition in wafer maps', *Procedia Computer Science*, Vol. 217, pp.570–583.

Wang, C.Y., Bochkovskiy, A. and Liao, H.Y.M. (2023) 'YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors', *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp.7464–7475.

Wang, J., Xu, C., Yang, Z. et al. (2020) 'Deformable convolutional networks for efficient mixed-type wafer defect pattern recognition', *IEEE Transactions on Semiconductor Manufacturing*, Vol. 33, No. 4, pp.587–596.

Wang, X., Jia, X., Jiang, C. et al. (2022) 'A wafer surface defect detection method built on generic object detection network', *Digital Signal Processing*, Vol. 130, p.103718.

Woo, S., Park, J., Lee, J.Y. et al. (2018) 'CBAM: convolutional block attention module', *Proceedings of the European Conference on Computer Vision (ECCV)*, pp.3–19.

Xia, B., Li, S., Li, N. et al. (2023) 'Surface defect recognition of solar panel based on percollation-based image processing and Serre standard model', *IEEE Access*, Vol. 11, pp.55126–55138.

Yang, J., Xu, Y., Rong, H.J. et al. (2020) 'A method for wafer defect detection using spatial feature points guided affine iterative closest point algorithm', *IEEE Access*, Vol. 8, pp.79056–79068.

Yu, C., Feng, Z., Wu, Z. et al. (2023) 'HB-YOLO: an improved YOLOv7 algorithm for dim-object tracking in satellite remote sensing videos', *Remote Sensing*, Vol. 15, No. 14, p.3551.

Yu, N., Xu, Q. and Wang, H. (2019) 'Wafer defect pattern recognition and analysis based on convolutional neural network', *IEEE Transactions on Semiconductor Manufacturing*, Vol. 32, No. 4, pp.566–573.

Yuan, J., Zhang, N., Xie, Y. et al. (2022) 'Detection of prohibited items based upon X-ray images and improved YOLOv7', *Journal of Physics: Conference Series*, IOP Publishing, Vol. 2390, No. 1, p.012114.

Zhao, Z., Wang, J., Tao, Q. et al. (2024) 'An unknown wafer surface defect detection approach based on Incremental Learning for reliability analysis', *Reliability Engineering & System Safety*, Vol. 244, p.109966.

Zheng, J. and Zhang, T. (2023) 'Wafer surface defect detection based on background subtraction and faster R-CNN', *Micromachines*, Vol. 14, No. 5, p.905.