Tile Defect Recognition Network Based on Amplified Attention Mechanism and Feature Fusion

JiaMing Zhang, ZanXia Qiang, YuGang Li
School of Computer Science, Zhongyuan University of Technology, Zhengzhou, China

***Abstract*—For the current situation of low AP of tile defect detection with incomplete detection of defect types, this paper proposes YOLO-SA, a detection neural network based on the enhanced attention mechanism and feature fusion. We propose an enhanced attention mechanism named amplified attention mechanism to reduce the information attenuation of the defect information in the neural network and improve the AP of the neural network. Then, we use the EIoU loss function, the four-layer feature fusion, and let the backbone network directly involved in the detection and other methods to construct an excellent tile defect detection and recognition model Yolo-SA. In the experiments, this neural network achieves better experimental results with an improvement of 8.15 percentage points over Yolov5s and 8.93 percentage points over Yolov8n. The model proposed in this paper has high application value in the direction of tile defect recognition.**

***Keywords*—Amplified attention mechanism; defect recognition; small target recognition; Yolo; feature fusion**

I. INTRODUCTION

Tile defect detection is an essential part of modern industrial production. Tiles are widely used in the manufacturing, construction and decoration industries for flooring, walls, kitchens and bathrooms. However, due to various factors in the production process, various defects can appear on the surface of tiles, such as cracks, unevenness, color variations and stains. The detection of these defects is still plagued by a large number of small targets, variable and irregular shapes, inconspicuous features and other factors, companies in the manufacturing process cannot avoid producing tiles with various types of defects. These defects not only affect the aesthetics, but can also lead to a decrease in the functionality and durability of the tiles. Therefore, tile surface defect detection is a key task in visual inspection, the goal of which is to automatically detect and recognise possible defects, damage or undesirable features on the tile surface. A good tile defect recognition model can help companies to improve quality, save manual inspection costs, increase productivity, reduce defect rate, reduce environmental impact and energy consumption.

Several advances have been made in the field of tile defect detection. Traditional methods are mainly based on image processing techniques such as edge detection, texture analysis and shape matching. However, these methods are difficult to detect defects in complex textured backgrounds. In recent years, the development of deep learning techniques has brought new opportunities for tile defect detection. CNN, Faster-RCNN [1], Yolov3, Yolov5 [2], etc. have average performance in defect detection and still need further improvement.

Y Huang et al. [3] implemented tile defect segmentation using MCue, U-Net [4] and Push networks by generating three channels of resized inputs with MCue, including an MCue saliency image and two original images; U-Net learns the most informative regions, which is essentially a deeply structured convolutional network; and Push network defines the prediction of defects through two fully connected layers and an output layer constructed to define the exact location of the predicted surface defects. The model can detect multiple surface defects from low-contrast images, but it cannot accurately detect multiple defects generated in real production. Wan G et al. [5] improved yolov5 by deepening the network layers and incorporating a Convolutional Block Attention Module (CBAM) [6] attention mechanism, and replaced the original convolution with a depth-separable convolution, obtaining a lightweight model that can detect small targets. Lu Q et al. [7] proposed an intelligent surface defect detection method for ceramic tiles based on the improved YOLOv5s algorithm, using Shufflenetv2 [8], Path Aggregation Network (PAN) [9], Feature Pyramid Network [10] and the attention mechanism to improve the model and achieve a lightweight and high-performance tile defect detection. Xie L et al. [11] proposed fusion feature CNN and added an attention mechanism to realise efficient tile surface defect detection. H Lu et al. [12] collected 1241 samples and realised tile defect detection based on acoustic waves and proposed a cross-attention mechanism based on acoustic wave information features to make the model defect detection, the final accuracy rate is 98.8%. Although the method is effective, the implementation cost of the method is relatively high. Stephen O et al. [13] used a hand-designed neural network to achieve the detection of cracks in floor tiles with an accuracy rate of 99.43%.

In this paper, methods such as Amplified Attention (AA) mechanism, 4-layer feature fusion, and direct addition of backbone network feature information to the detection header are proposed to improve the performance of the model. AA mechanism is a method used to improve the performance of the model. In tile defect detection, the AA mechanism can help the model to pay more attention to the defective region, thus improving the detection accuracy. The specific process is that by introducing the cross-channel attention mechanism into the convolutional neural network, the model can better capture the characteristics of the defects. This approach can further improve the performance of tile defect detection. Feature fusion improves the overall amount of feature information

captured by the model, which in turn improves the performance of the tile defect detection model. To enable the detection head to acquire more feature information, feature information from the backbone network was added to the detection head in this study. Finally, the Efficient Intersection over Union (EIoU) loss function was used to improve the accuracy of the model in predicting the direction of movement of the frame and to improve the accuracy of the model in predicting defects.

II. DATASETS

In this paper, we use the tile defect detection dataset provided by the Guangdong Province Tile Defect Detection Competition of 2021 Ali Tianchi, which consists of 5,388 images with image resolutions of 8192px × 6,000px and 4096px × 3,500px. The dataset is labelled with six categories, namely Edge exception, Angular exception, White dotted flaw, Light color block flaw, Dark dotted flaw and Aperture flaw. Examples of ceramic tile defects are shown in Fig. 1.

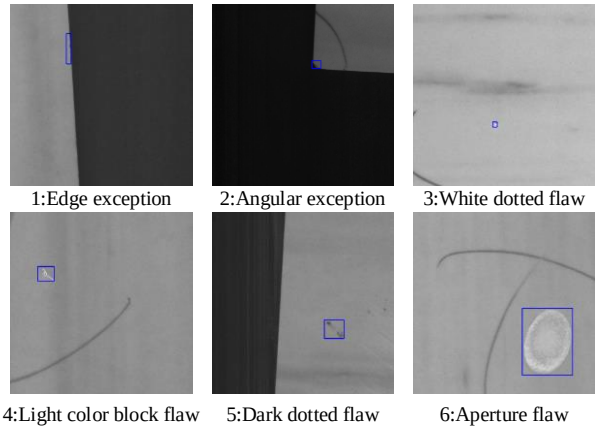

Fig. 1. Example of tile defects.

A. Image Segmentation

The image labelling frame statistics are shown in Fig. 2A, from which it can be seen that the labelling frames are mostly concentrated between 0 and 0.05, the size of the labelling frames for tile defects is extremely small relative to the whole image. If the whole image is fed into the neural network during training, the training speed will be very slow. Therefore, here the image is segmented for processing, the image segmentation method is that each image contains at least one tile defect detection point location, if it does not contain a tile defect detection point location, the segmented portion is considered as an invalid portion, and this segmented image is not generated. According to this method, the original image is segmented into images of 640 pixels × 640 pixels and a total of 19960 images are obtained. The statistics of the labelled boxes after segmentation are shown in Fig. 2B, from which it can be seen that the size of the labelled boxes of the segmented tile defects with respect to the whole image is significantly improved compared to the pre-segmentation.

B. Data Augmentation

Models built from datasets with sufficient amount of data have stronger robustness and generalisation ability. Therefore, in order to improve the performance of the tile defect detection model, online augmentation of the segmented tile defect dataset is improved. Online augmentation is the augmentation of the dataset during the training process, and in each epoch, a random augmentation is performed in proportion to the set augmentation strategy. The data augmentation strategies are HSV augmentation, flip, mosaic, zoom and pan, and the augmentation ratio of these data augmentation strategies in each epoch is 0.15, 0.5, 0.1, 0.5 and 0.3, respectively. The effect of the image after augmentation is shown in Fig. 3.

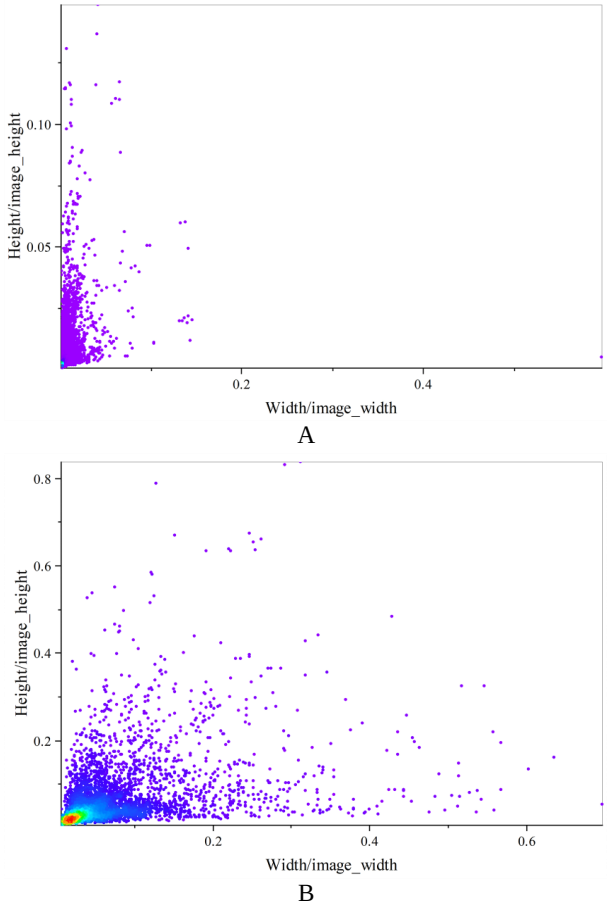

Fig. 2. Comparison of the statistics of defective labeling frames of tiles before segmentation and after segmentation.

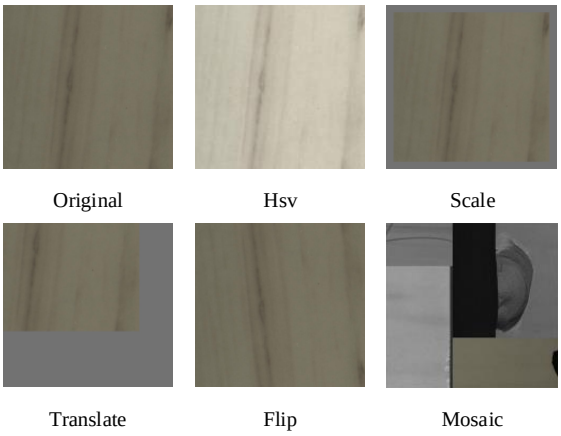

Fig. 3. Dataset augmentation.

III. Experimental Design and Methodology

Firstly, a general overview of the YOLO-SA target detection neural network model is given, and then the design of the neural network structure of the YOLO-SA model is discussed from the three parts of the backbone network, the neck network and the detection head, respectively. In the YOLO-SA neural network structure, the backbone network is responsible for extracting the feature information of the tile defect image, the neck network can fuse the shallow feature information extracted by the backbone network with the deep feature information to improve the feature information extraction ability of the neural network, and the detection head detects the feature information obtained from the neck network. The YOLO-SA network structure is shown in Fig. 4. In Fig. 4, conv, BN, LRelu, Silu [14], avgpool, maxpool denote convolu-tional computation, batch normalisation, leaky-Relu activation function, Silu activation function, average pooling and maximum pooling, respectively, and concat denotes the channel connection operation.

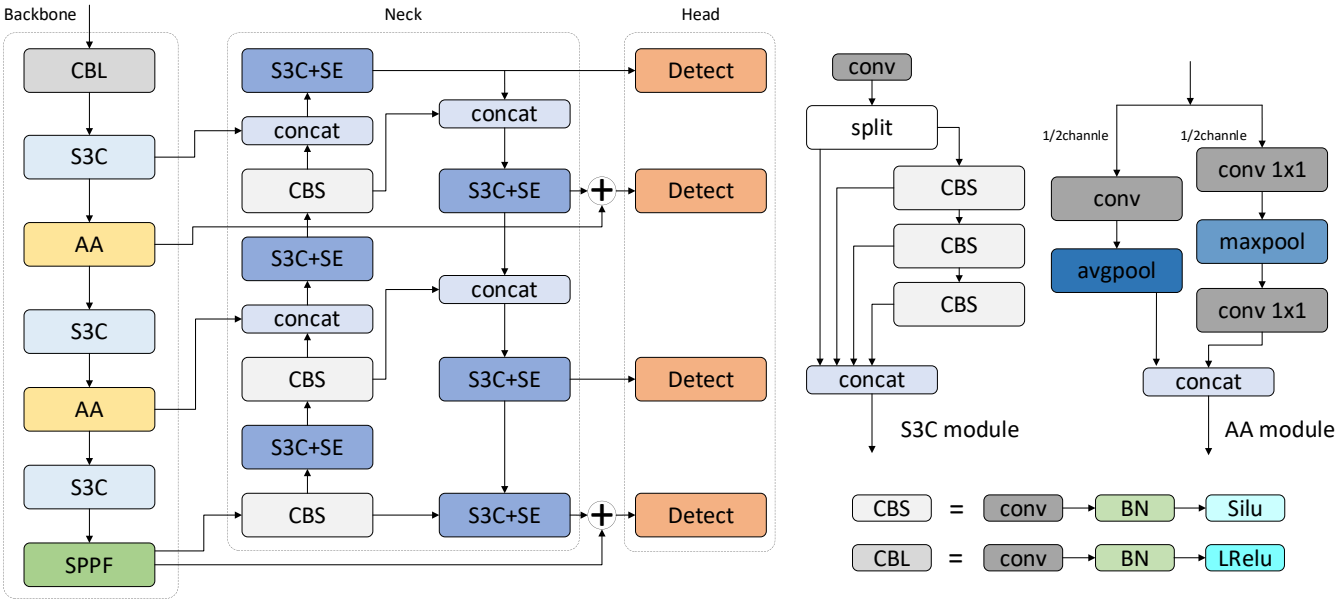

Fig. 4. YOLO-SA Neural network structure.

A. Overview

In YOLO-SA, backbone network includes CBL module, S3C module, AA module and Spatial Pyramid Pooling Fast (SPPF) [15] module, CBL module can effectively extract flat feature information, S3C module has very powerful feature information extraction ability, and AA module can effectively reduce degree of information loss in backbone network. Compared with the traditional CNN-based backbone feature extraction network, the backbone network of YOLO-SA only uses the information provided by the region when obtaining the target feature information, and this backbone network has a global modelling capability and a powerful remote dependency, which can better detect tile defects.

The SPPF module can fully extract the deep feature information from the backbone network, and the structure of the SPPF module is shown in Fig. 5. The main function of the SPPF is to perform a convolution operation on each region before the pooling operation, and to combine the convolution result and the pooling result as the output features. This method can retain more local feature information and improve the accuracy of the network. The appearance of SPPF makes the network more adaptable to objects of different sizes and effectively avoids problems such as image distortion caused by cropping and scaling operations of image regions. The calculation formula is:

$$\begin{aligned} AA = concat([F, maxpool(F), \\ maxpool(maxpool(F)), \\ maxpool(maxpool(maxpool(F)))], 1) \end{aligned} \quad (1)$$

where, F denotes the input feature map.

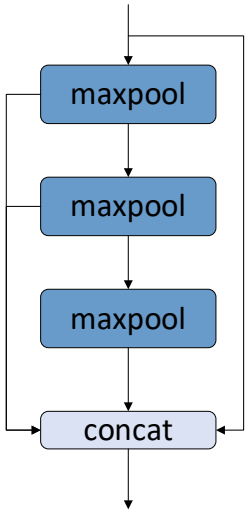

Fig. 5. SPPF module.

In the neck network of YOLO-SA, the original three-layer PANet feature pyramid is expanded to four layers to fully integrate the feature information extracted from the backbone network. In addition, the Squeeze-and-Excitation Module (SE) [16] attention mechanism is added to this neck network to increase the attention to the target information in the spatial dimension and further improve the performance of the model. The SE has excellent information extraction ability, and at the same time, this attention mechanism requires much less computation compared to the CA mechanism, the CBAM, and so on. Therefore, the SE is used in YOLO-SA. The structure of the SPPF module is shown in Fig. 6.

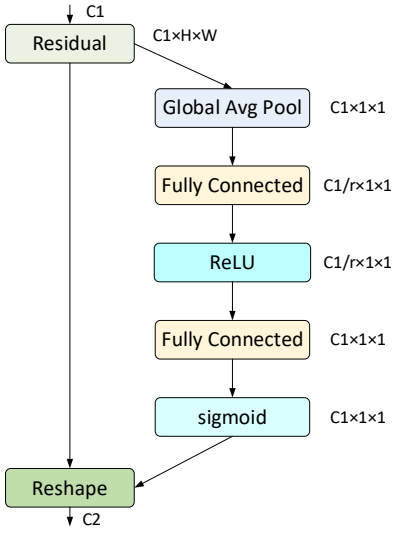

Fig. 6. SE module.

B. Loss Functions

1) IoU: Intersection over Union (IoU), which is calculated as:

$$IoU = \frac{|A \cap B|}{|A \cup B|} \tag{2}$$

This means that the intersection of two regions is more than the concatenation of the last two sets. The visual representation is shown in Fig. 7.

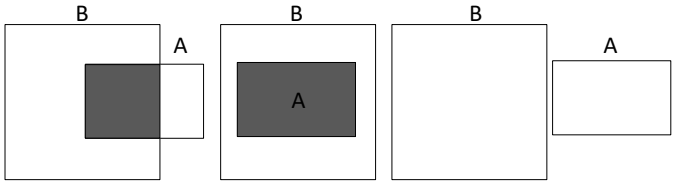

Fig. 7. IoU concept

2) GIoU, DIoU, CIoU: Since IOU is calculated only for the overlapping region between the predicted and real frames and does not focus on the non-overlapping region, H Rezatofighi et al. [17] developed the Generalized Intersection over Union (GIOU) loss calculation function, which is formulated as:

$$GIoU = IoU - \frac{|C-U|}{|C|}, U = A \cup B \tag{3}$$

where C denotes the area of the minimum closure area of the prediction frame and the real frame, and U is the area of the concatenation of the prediction frame and the real frame. The image representation of each parameter is shown in Fig. 8.

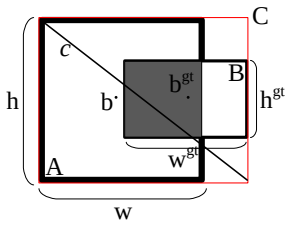

Fig. 8. Explanation of each parameter of the GIoU, DIoU and CIoU

In order to obtain more information to better represent the gap between the prediction box and the real box, Zhaohui Zheng et al. proposed Distance Intersection over Union (DIoU) [18], DIoU adds more information into the regression calculation, such as the distance between the prediction box and the real box, and the size of the prediction box and the real box. The formula of DIoU is:

$$DIoU = IoU - \frac{\rho^2(b,b^{gt})}{c^2} \tag{4}$$

where b, b^{gt} represent the centroids of the predicted and real images, respectively, $\rho(.)$ represents the computed Euclidean distance, and c represents the diagonal distance of the minimum closure region. The image representation of each parameter is shown in Fig. 8.

The factors considered by DIoU are still not able to meet the needs of loss calculation in practice. DIoU does not measure the difference in the size of the predicted frame and the real frame, so Zhaohui Zheng et al. [18] proposed Complete Intersection over Union (CIoU), which is calculated by the formula:

$$\begin{gathered} CIoU = DIoU - \alpha v \\ \alpha = \frac{v}{(1-IoU)+v} \\ v = \frac{4}{\pi^2}(arctan\frac{w^{gt}}{h^{gt}} - arctan\frac{w}{h})^2 \end{gathered} \tag{5}$$

where α is the weight function, v is used to measure the similarity of the width-to-height ratio of the predicted frame to the real frame, and w, h, w^{gt}, and h^{gt} denote the width of the predicted frame, the height of the predicted frame, the width of the real frame, and the height of the real frame, respectively. The picture explanation is shown in Fig. 8.

CIoU adds the detection frame scale loss to DIoU, which allows the prediction frame to more accurately match the real frame by taking into account the length and width loss. The CIoU loss ($L_{CIoU} = 1 - CIoU$) can help the model to converge accurately and quickly during training, and to predict targets in complex backgrounds more accurately.

3) EIoU: The most important thing in YOLO-SA's recognition head is the loss function. The purpose of the loss function is mainly to make the model localisation more accurate and the recognition accuracy higher. In the process of tile defect recognition, because the tile defect target is very

small, in order to accurately recognise the feature information, so the box loss in YOLO-SA uses a more advanced EIoU loss [19], which can more accurately measure the difference between the predicted bounding box and the real bounding box. The EIoU loss is calculated as:

$$EIoU = IOU - \frac{\rho^2(\boldsymbol{b},\boldsymbol{b^{gt}})}{(w^c)^2+(h^c)^2} - \frac{\rho^2(w,w^{gt})}{(w^c)^2} - \frac{\rho^2(h,h^{gt})}{(h^c)^2} \quad (6)$$

where w^c and h^c denote the width and height of the minimum closure region, respectively. the EIoU parameter image is explained as shown in Fig. 9.

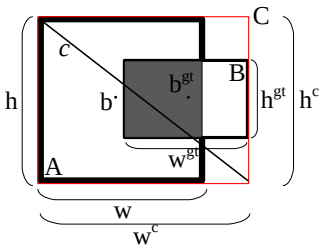

Fig. 9. Explanation of each parameter of the EIoU

4) Classification loss and box loss: The classification loss function L_{cls} used in the YOLO-SA network is formulated as:

$$\mathrm{y_i} = \mathrm{Sigmoid}(\mathrm{x_i}) = \frac{1}{1+\mathrm{e}^{-\mathrm{x_i}}} \quad (7)$$

$$BCE = -\sum_{\mathrm{n=1}}^{\mathrm{N}} \mathrm{y_i^*} \log(\mathrm{y_i}) + (1-\mathrm{y_i^*}) \log(1-\mathrm{y_i}) \quad (8)$$

$$L_{cls}(\boldsymbol{c}_{\mathrm{p}}, \boldsymbol{c}_{\mathrm{gt}}) = BCE_{cls}^{sig}(\boldsymbol{c}_p, \boldsymbol{c}_{gt}; w_{cls}) \quad (9)$$

where N is the total number of categories, $\mathrm{x_i}$ is the predicted value of the current category, $\mathrm{y_i}$ is the probability of the current category obtained according to the activation function, and $\mathrm{y_i^*}$ is the true value of the current category (0 or 1), $\boldsymbol{c}_p$ is the predicted probability of the category, $\boldsymbol{c}_{gt}$ is the ground truth of the category, and w_{cls} is the weight of the current category. The confidence loss function is:

$$L_{obj}(p_0, p_{iou}) = BCE_{obj}^{sig}(p_0, p_{iou}; w_{obj}) \quad (10)$$

where p_0 is the confidence score of the target, p_{iou} is the iou value of the prediction frame and the corresponding target frame, and w_{obj} is the current target weight.

C. S3C Module

Compared with the Transformer module, the S3C module can effectively reduce the amount of computation and hardware requirements, and at the same time has almost the same ability to extract image information as the Transformer module. As shown in Fig. 1, the S3C module first splits the input channel, part of which is directly involved in the concatenation calculation, and the other part is calculated three times by the CBS module, and the result is calculated by the concatenation calculation after each calculation, and the final calculation result is obtained after the concatenation calculation is finished. After the experiment, it is proved that the module has excellent performance in the tile defect dataset. The calculation formula of S3C module is:

$$\begin{aligned} S3C = concat([&split, CBS(split),\\ &CBS(CBS(split)),\\ &CBS(CBS(CBS(split))],1) \end{aligned} \quad (11)$$

D. Amplified Attention Mechanism

The AA mechanism is proposed to address the situation that the tile defect target in this dataset is small and difficult to detect accurately. The structure of the enhanced attention mechanism is shown in Fig. 1. The avgpool on the left can obtain hierarchical feature information, which can better distinguish the target area from the non-target area. The $conv_{1\times1}$ structure on the right can further deepen the feature information obtained by the higher-level network; the maxpool can highlight the feature information of the deep feature map, thus further highlighting the target region; the $conv_{1\times1}$ structure at the back can narrow the depth of the image to facilitate the concatenation operation. The enhanced attention mechanism can reduce the degree of feature information loss during neural network training. The formula for the AA module is:

$$\begin{aligned} AA = concat([&avgpool(conv),\\ &conv_{1\times1}(maxpool(conv_{1\times1}))],1) \end{aligned} \quad (12)$$

E. Four-Layer Feature Information Fusion

To further improve the performance of the neural network model for tile defect detection, a 4-layer PANet fusion module is proposed and the corresponding detection head is added to this module. The feature information fusion module is shown in the neck part of Fig. 1. In addition, to make the information flow more appropriate and reduce the loss of feature information, the computation results of an AA module and SPPF module in the backbone network are directly fused with the 2nd and 4th detection heads before operation.

IV. Experimental Environment and Evaluation Indicators

A. Experimental Environment

Experimental platform: OS Windows 11, CPU i9-12900K, GPU RTX5000 24GB, RAM 64GB, Pytorch 2.0.1, CUDA 11.8, PyCharm 2022.2.1, Anaconda 22.11.1.

In this study, the segmented dataset is divided into three parts according to the ratio of 8:1:1, which are training set, validation set and test set. The input size of the neural network for the tile defect image dataset is 640 pixels × 640 pixels. The optimiser uses AdamW [20] with momentum set to 0.9, an initial learning rate of 0.001, and 100 iterations, keeping only the optimal model and the model produced by the last iteration.

B. Evaluation Indicators

The evaluation indicators used in this study include Precision, Recall, and mAP@0.5. Precision represents how many of the predicted positive samples are truly positive samples, Recall represents how many of the positive examples in the sample were predicted correctly, and mAP@0.5 represents the average accuracy of m categories when IoU is 0.5. The calculation formulas of precision and recall are shown in (1) and (2); the calculation formula of mAP@0.5 is shown in formula (3).

$$Precision = \frac{TP}{TP+FP} \quad (13)$$

$$Recall = \frac{TP}{TP+FN} \quad (14)$$

where TP is the number of positive classes predicted as positive, FP is the number of negative classes predicted as negative and FN is the number of negative classes predicted as negative.

$$mAP@0.5 = \frac{1}{m}\sum_{i\in m}^{m}\int_0^1 P(r_i)dr_i \quad \text{IoU} = 0.50 \quad (15)$$

$P(r_i)$ denotes the correspondence between recall and precision; mAP@a denotes the average precision of m categories when IoU is a.

V. Experimental Results and Analysis

A. Performance Comparison Before and After Image Segmentation

Image segmentation has a great impact on the performance of the model, and segmentation can increase the size of the target and make it easier to detect. The performance comparison before and after segmentation on Yolov8n is shown in Table I. From the table, we can see that image segmentation has significantly improved the performance of the model.

TABLE I. Performance Comparison Before and After Segmentation on Yolov8n

Image segmentation	Precision	Recall	mAP@0.5
$Before_{640\times640}$	0.5712	0.1396	0.1727
$Before_{1280\times1280}$	0.4027	0.1954	0.2281
$After_{640\times640}$	0.5206	0.6813	0.6133

B. Performance Comparison Using the EIoU Loss Function

The comparison of the effectiveness of GIoU, CIoU and EIoU is shown in Table III. The CIoU loss function uses proportions to determine whether the size of the prediction frame is met or not, as there are very many small targets in this dataset, it is not easy to determine the prediction frame in terms of width to height proportions, so the EIoU loss function achieves a better result among these three loss functions.

TABLE II. Comparison of the Effects of EIoU, CIoU and GIoU on Yolov8n

Loss Function	Precision	Recall	mAP@0.5
GIoU	0.3193	0.3724	0.3865
CIoU	0.5206	0.6813	0.6133
EIoU	0.5432	0.6859	0.6241

C. Performance Comparison between 4-layer Feature Fusion Module and 3-layer Feature Fusion Module

The 4-layer feature fusion module can obtain more deep information, which can enable the model to detect defects more accurately, and its performance is shown in Table IV.

D. Performance Comparison of the Enhanced Attention Mechanism with other Attention Mechanisms

To verify the effect in the tile recognition dataset, a comparison experiment is designed here to verify the effect comparison with other attention mechanisms, as shown in Table II. From Table II, we can see that the effect of AA attention is better than that of other attention mechanisms. A From Table II, we can see that the improvement of tile defect detection is better than that of other attention mechanisms, thanks to the feature of the AA mechanism of reducing the loss of feature information during the training process of the neural network.

TABLE III. Performance Comparison of Feature Fusion Modules with Different Number of Layers on Yolov8n

Number of layers in the fusion part	Precision	Recall	mAP@0.5
Three-layer feature fusion	0.5206	0.6813	0.6133
Four-layer feature fusion	0.5379	0.6935	0.6472

TABLE IV. Performance Comparison of AA Mechanism with Other Attention Mechanisms in Yolo-SA

Module name	Precision	Recall	mAP@0.5
SE	0.5820	0.7351	0.6925
CA	0.5783	0.7291	0.6892
ECA	0.5769	0.7274	0.6744
CBAM	0.5838	0.7418	0.6892
Muti-Head Attention	0.5981	0.7353	0.7011
AA	0.6032	0.7527	0.7024

E. Feature Information in the Backbone Network Added to the Detection Head

From Table V, it can be seen that the detection results after the backbone network is added to the detection head are improved over the detection results before it is added, and the method can effectively improve the performance of the model.

TABLE V. Comparison of the Effect of Yolo-SA Backbone Network Feature Information Before and After Adding it Directly to the Detection Header

Method	Precision	Recall	mAP@0.5
	0.5206	0.6813	0.6133
AA	0.5515	0.7381	0.6713
SPPF	0.5729	0.7442	0.6739
AA+SPPF	0.6032	0.7527	0.7024

VI. Conclusion

The mAP@0.5 curves and loss functions of Yolov5s, Yolov8n, and Yolo-SA are shown in Fig. 10. Yolo-SA outperforms Yolov5s and Yolov8n, proving that the Yolo-SA model has a good ability to detect tile defects.

First, we dramatically improve the accuracy of defect detection by using image segmentation techniques, and the defect detection mAP@0.5 after segmentation is 48 percentage points higher than before segmentation at the same input

resolution. Then, we use the EIoU loss function, AA attention mechanism, four-layer feature fusion, and let the backbone network directly participate in the detection to construct an excellent tiled defect detection and recognition model, which is capable of recognizing and detecting multiple defects in multiple complex backgrounds. The mAP@0.5 of the Yolo-SA model improves by 8.15 percentage points and 8.93 percentage points compared with that of the Yolov5s and the Yolov8n, respectively. The mAP@0.5 of the Yolo-SA model is improved by 8.15 percentage points and 8.93 percentage points compared to that of the Yolov5s and Yolov8n, respectively. The Yolo-SA model is able to detect tile defects under a variety of environments, and the actual detection results are shown in Fig. 11.

At present, the performance of the Yolo-SA model still has a lot of room for optimization. In practical applications, the Yolo-SA model can only be used as an auxiliary model for artificial tile defect detection. In the future, large models can be combined to further improve the performance of the ceramic tile defect detection model, which can further improve the accuracy of ceramic tile defect detection in actual production scenarios, reduce unnecessary production, and thereby reduce energy consumption and environmental pollution.

TABLE VI. PERFORMANCE OF YOLOV5S, YOLOV8N, AND YOLO-SA

Model	Precision	Recall	mAP@0.5
Yolov5s	0.5310	0.7222	0.6209
Yolov8n	0.5130	0.6800	0.6131
Yolo-SA	0.6032	0.7527	0.7024

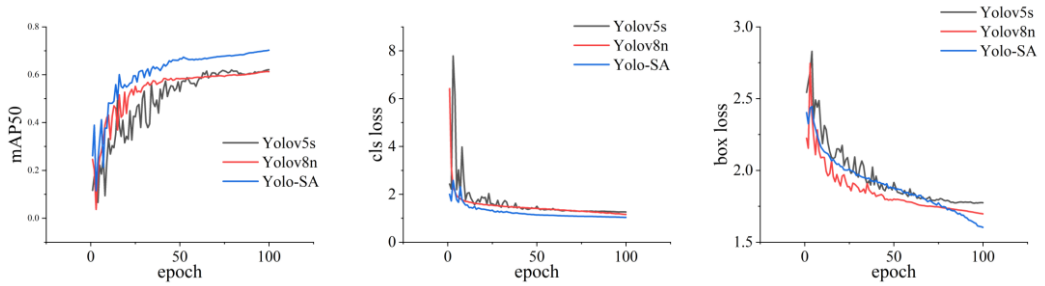

Fig. 10. mAP@0.5 curve, cls loss curve, box loss curve.

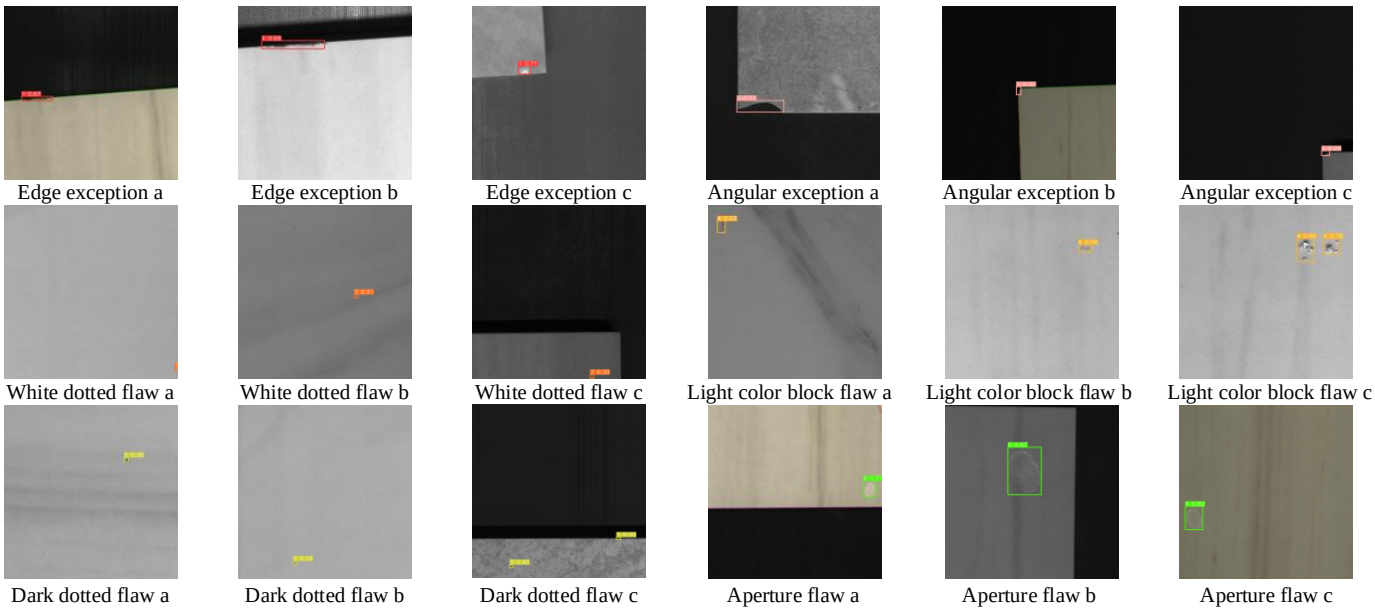

Fig. 11. Yolo SA detection results.

REFERENCES

[1] Girshick, R. (2015). Fast r-cnn. In Proceedings of the IEEE international conference on computer vision (pp. 1440-1448).

[2] Jiang, P., Ergu, D., Liu, F., Cai, Y., & Ma, B. (2022). A Review of Yolo algorithm developments. Procedia computer science, 199, 1066-1073.

[3] Huang, Y., Qiu, C., & Yuan, K. (2020). Surface defect saliency of magnetic tile. The Visual Computer, 36(1), 85-96.

[4] Du, G., Cao, X., Liang, J., Chen, X., & Zhan, Y. (2020). Medical Image Segmentation based on U-Net: A Review. Journal of Imaging Science & Technology, 64(2).

[5] Wan, G., Fang, H., Wang, D., Yan, J., & Xie, B. (2022). Ceramic tile surface defect detection based on deep learning. Ceramics International, 48(8), 11085-11093.

[6] Woo, S., Park, J., Lee, J. Y., & Kweon, I. S. (2018). Cbam: Convolutional block attention module. In Proceedings of the European conference on computer vision (ECCV) (pp. 3-19).

[7] Lu, Q., Lin, J., Luo, L., Zhang, Y., & Zhu, W. (2022). A supervised approach for automated surface defect detection in ceramic tile quality control. Advanced Engineering Informatics, 53, 101692.

[8] Ma, N., Zhang, X., Zheng, H. T., & Sun, J. (2018). Shufflenet v2: Practical guidelines for efficient cnn architecture design. In Proceedings of the European conference on computer vision (ECCV) (pp. 116-131).

[9] Liu, S., Qi, L., Qin, H., Shi, J., & Jia, J. (2018). Path aggregation network for instance segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 8759-8768).

[10] Lin, T. Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2017). Feature pyramid networks for object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2117-2125).

[11] Xie, L., Xiang, X., Xu, H., Wang, L., Lin, L., & Yin, G. (2020). FFCNN: A deep neural network for surface defect detection of magnetic tile. IEEE Transactions on Industrial Electronics, 68(4), 3506-3516.

[12] Lu H, Zhu Y, Yin M, et al. Multimodal fusion convolutional neural network with cross-attention mechanism for internal defect detection of magnetic tile[J]. IEEE Access, 2022, 10: 60876-60886.

[13] Stephen, O., Maduh, U. J., & Sain, M. (2021). A machine learning method for detection of surface defects on ceramic tiles using convolutional neural networks. Electronics, 11(1), 55.

[14] Ramachandran, P., Zoph, B., & Le, Q. V. (2017). Searching for activation functions. arXiv preprint arXiv:1710.05941.

[15] He, K., Zhang, X., Ren, S., & Sun, J. (2015). Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE transactions on pattern analysis and machine intelligence, 37(9), 1904-1916.

[16] Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-excitation networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 7132-7141).

[17] Rezatofighi, H., Tsoi, N., Gwak, J., Sadeghian, A., Reid, I., & Savarese, S. (2019). Generalized intersection over union: A metric and a loss for bounding box regression. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 658-666).

[18] Zheng, Z., Wang, P., Liu, W., Li, J., Ye, R., & Ren, D. (2020, April). Distance-IoU loss: Faster and better learning for bounding box regression. In Proceedings of the AAAI conference on artificial intelligence (Vol. 34, No. 07, pp. 12993-13000).

[19] Zhang, Y. F., Ren, W., Zhang, Z., Jia, Z., Wang, L., & Tan, T. (2022). Focal and efficient IOU loss for accurate bounding box regression. Neurocomputing, 506, 146-157.

[20] Loshchilov, I., & Hutter, F. (2017). Decoupled weight decay regularization. arXiv preprint arXiv:1711.05101.