Target Detection of Leakage Bubbles in Stainless Steel Welded Pipe Gas Airtightness Experiments Based on YOLOv8-BGA

Huaishu Hou, Zikang Chen*, Chaofei Jiao
School of Mechanical Engineering, Shanghai Institute of Technology, Shanghai, 201418, China

***Abstract*—Gas-tightness experiment is an effective means to detect leakage of stainless steel welded pipe, and the vision-based bubble recognition algorithm can effectively improve the efficiency of gas-tightness detection. This study proposed a new detection network of YOLOv8-BGA using the YOLOv8 model as a baseline, which can achieve effective identification of leakage bubbles and bubble images are collected under different lighting conditions in a practical industrial inspection environment to create a bubble dataset. Firstly, a C2f_BoT module was designed to replace the C2f module in the backbone network, which improved the feature extraction capability of the model; secondly, the convolutional layer of the neck network was replaced by using the GSConv module, which achieved the model lightweighting; thirdly, the C2f-BM attention mechanism was added before the detection layer, which effectively improved the model performance; finally, the WIoU was used to improve the loss function, which improved the detrimental effect of small bubbles of low-quality samples in the dataset on the gradient, and significantly improved the convergence speed of the network. The experimental results showed that the average leakage bubble detection accuracy of the YOLOv8-BGA model algorithm reached 97.7%, which improved by 5.3% compared with the baseline, and meets the needs of practical industrial inspection.**

Keyword—Image processing; stainless steel welded pipe; non-destructive testing; YOLOv8; attention mechanism; loss function

I. INTRODUCTION

Stainless steel welded pipe has a wide range of uses, and is widely used in chemical industry, automobile manufacturing, shipbuilding and other industrial production areas [1], but it is prone to leakage problems, which has a greater impact on the subsequent use of the product. Gas tightness test is an effective means to test whether the stainless steel welded pipe leakage, when the welded pipe leakage exists, air bubbles are generated in the water, and the identification of air bubbles is an important indicator for judging the gas tightness. The traditional airtightness detection for workers to observe, the method is subject to subjective factors, easy to miss, low detection efficiency [2-4].Vision-based airtightness testing can exclude the influence of human subjective factors, assess the results through the indicators, make the test results more accurate, and improve the efficiency of product testing. Vision-based airtightness detection has traditional image processing method [5] and deep learning target detection based method [6].

Traditional image processing algorithms rely on classical image processing techniques based on edge contour and circumferential curvature fitting to manually extract the bubble edge features. Qaddoori [7] used the use of Hough's Circle Transform algorithm to identify the tiny bubbles in the graph based on Canny operator and segmented the centre and edges of the bubbles by two thresholds to calculate the average diameter and number of bubbles in the graph; Wen [8] designed a new image processing algorithm based on the concept of differential segmentation while considering the geometry and deflection angle of the bubbles; Akdemir [9] proposed a detection method using wavelet transform denoising and entropy threshold segmentation. The above proposed methods need to modify the parameters to respond to different detection environments, and have weak generalisation ability for identifying bubbles, and are prone to miss detection and false detection. Deep learning based target detection method has higher detection rate, stronger generalisation ability and better robustness for bubble recognition.

Deep learning target detection algorithms are classified into two types: a two-stage algorithm with high accuracy but slow speed, such as Faster R-CNN [10], Mask R-CNN [11]; and a single-stage algorithm with a simple structure and high computational efficiency, such as Krysko, N.V.et al [12] through the integrated application of non-destructive testing techniques, computer vision and convolutional neural networks, the surface of pipeline pitting and defects were classified and quantitatively analysed; Zhao et al [13] proposed a 3D quartz crucible bubble detection method, which significantly improved the detection accuracy of tiny bubbles by optimising the YOLOv5 network structure, introducing dilated convolution and ECA-Net mechanism, and combining Kalman filtering with Hungarian matching algorithm. Due to the different generalisation ability of the model in different detection environments. The real-time identification of bubbles is affected by the ripples generated by the bubble movement, the lighting environment, impurities in the water, sedimentation and other factors, which makes the collected bubble images have a lot of noise and missing bubble boundaries, which in turn affects the accuracy of target detection.

In this study, a model algorithm YOLOv8-BGA based on YOLOv8 is developed to achieve effective identification of leakage bubbles. Firstly, a C2f_BoT module is designed to replace the C2f module in the backbone network, which improves the feature extraction capability of the model; secondly, the convolutional layer of the neck network is replaced by using the GSConv module, which realises the model lightweight; again, the C2f-BM attention mechanism is

*Corresponding Author.

added before the detection layer, which effectively improves the model performance; and lastly, the loss function is improved by using the WIoU to improve the dataset The unfavourable effect of small bubbles of low-quality samples on the gradient significantly improves the convergence speed of the network.

II. Related Work

The first single-stage target detection YOLO algorithm was first proposed by Redmon [14] in 2015.The algorithm achieves fast identification and precise positioning of the target to be detected through a regression method. Currently, the YOLO series of algorithms has been developed to YOLOv10, and a variety of improved algorithms have been derived in academia. For example, Li [15] and others improved the internal structure of YOLOv5s and proposed a 'YOLOv5s-ShuffleNetV2-DWconv-Add' model, which provides an efficient acquisition method for fruit picking robots; Sun [16] and others proposed a 'Pconv-Wide lightweight' model for fruit picking robots. Pconv-Wide lightweight convolutional simplified YOLOv7 model, which increases the detection accuracy of UAVs on small targets while reducing the number of model parameters; Zhao [17] et al. proposed a Res-Clo network for denoising pre-processing of SAR images, and designed a DML-YOLOv8w network, which improves the performance of the model in multi-scale detection; YOLOv8 uses an Anchor-Free detection head [18], which improves the generalisation ability while reducing the model complexity, and improves the detection speed and accuracy compared to the previous YOLO algorithm.

YOLOv8 has four network model structures, YOLOv8n, YOLOv8s, YOLOv8m, and YOLOv8l, and the choice of the structure depends on the actual detection scenario and resource conditions. The YOLOv8n model is mainly divided into four parts: (1) Data enhancement strategies are applied to the image in the input part, such as Mosaic data enhancement and adaptive anchor frame computation to enhance the data diversity; (2) The backbone network is processed by multiple convolutional downsampling at key nodes, thus effectively reducing the size of the feature map and preserving the spatial information of the image. Then C2f modules of different sizes are used to capture multi-scale features; (3) The neck network accepts feature maps of various scales obtained through 8, 16, and 32 times downsampling from the backbone network, and fuses shallow high-resolution features and deep low-resolution features through PANet (path aggregation network), followed by fusing the fused feature maps with rich spatial and semantic information. Then the fused feature maps with rich spatial and semantic information are output to the Head part for further processing; (4) The Head network receives the multi-scale feature maps from Neck, performs up-sampling and splicing operations on them to match the resolution of the input image, and adjusts the number of channels through the feature conversion layer.

YOLOv8 has three Detect layers of different scales, 80×80, 40×40, and 20×20, which are used to detect large, medium, and small sized targets in the image, respectively. The three prediction layers output the category probability and bounding box location of each cell, and the stability of the prediction box is evaluated by a confidence layer.YOLOv8 adopts a Decoupled-Head structure to separate the classification and detection tasks to improve the detection efficiency. Finally, IOU (Intersection over union) [19] is used as the loss function for the bounding box regression, and then NMS (Non-Maximum Suppression) [20] removes the redundant prediction frames so as to avoid the same target from being detected repeatedly and retain the best results. This paper is based on the YOLOv8n algorithm to improve. The structure of the YOLOv8 model is shown in Fig. 1.

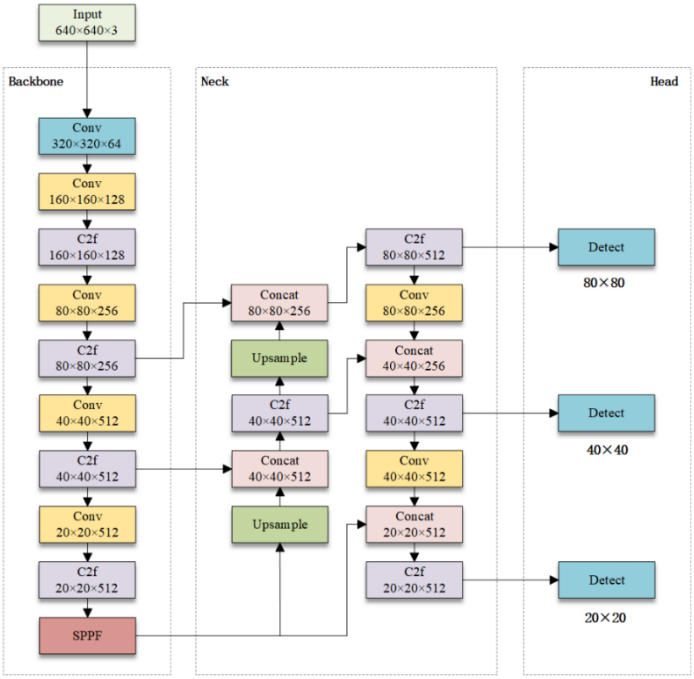


Fig. 1. YOLOv8 model structure.

III. Algorithm Design

A. Network Architecture Design

In this paper, four improvement modules for the YOLOv8n model are proposed. Firstly, the C2f_BoT module was constructed to replace the original C2f module of 40×40 and 20×20 sizes in the Backbone section, which improves the ability of the model to detect small and medium-sized bubbles; Secondly, the original Conv(Convolution) block is replaced by the GSConv in the neck network, which reduces the complexity of the model and improves the inference speed to achieve the effect of model lightweighting; Next, the C2f-BM attention mechanism is embedded in front of each detection head, which improves the detection accuracy of the model for small bubble targets; Finally, the WIoU loss function is introduced to speed up the network convergence. These improvements greatly improved the detection accuracy and detection speed of the original model in the stainless steel welded pipe gas-tightness inspection task. Fig. 2 shows the structure of the YOLOv8-BGA model obtained after the improvement.

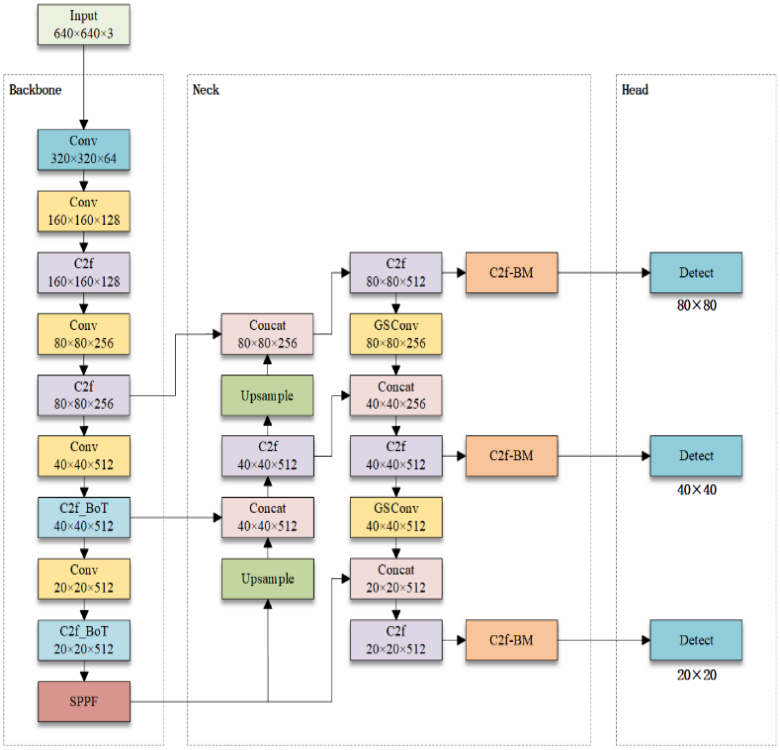
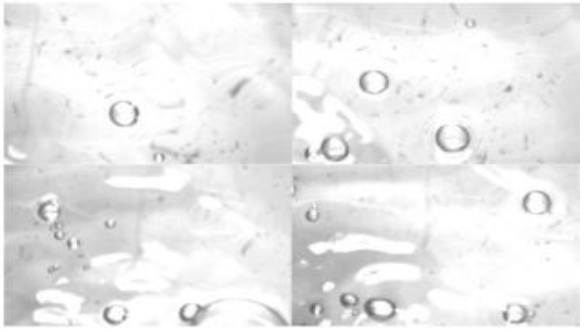


Fig. 2. YOLOv8-BGA model structure.

B. Mosaic Image Data Enhancement

Data enhancement of the self-constructed bubble dataset is performed using the Mosaic method at the YOLOv8n input. The Mosaic data enhancement algorithm improves the detection ability of the model in a small field of view by fusing multiple images into a single image according to a certain random scale. As shown in Fig. 3, four bubble images are randomly selected, certain parameters are set, and they are stitched into one image by random permutation, random size scaling, and random cropping.



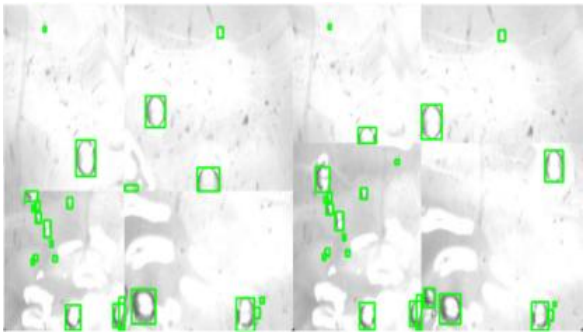


Fig. 3. Data enhancement.

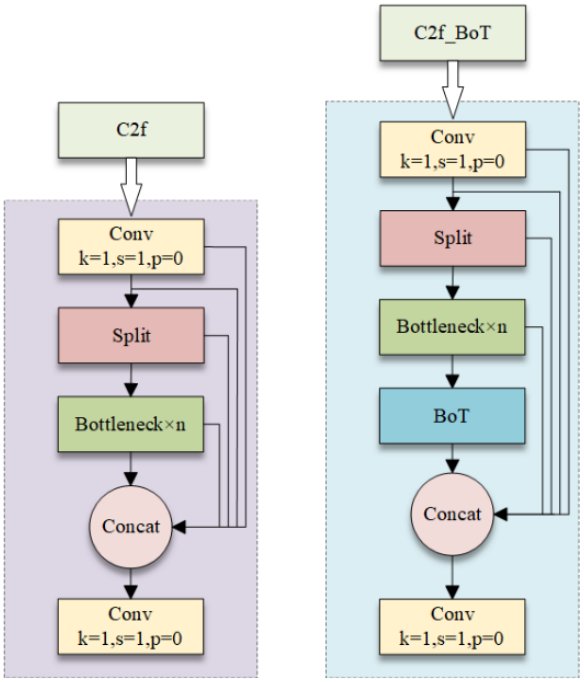
Using Mosaic data enhancement technique to strengthen the dataset can increase the data diversity, and the combined images obtained are more than the number of original images, which can get more small targets, and improve the generalisation ability of the trained model, which is a great improvement for detecting some tiny bubbles produced by subtle defects. And the network is able to batch process four image data in a single batch, which optimises the effect of the batch normalisation layer, thus reducing the computational burden on the GPU,

making it possible to achieve efficient training results on a single GPU.

C. C2f Module Improvements

The YOLOv8 network structure makes several uses of the C2f module [21], which is an improved residual block that enables feature fusion by connecting feature maps of different depths through splicing and upsampling operations. The spliced feature map fuses multi-scale features, which is helpful for the model to detect targets of different sizes. Compared with the traditional C3 module in YOLOv5, the C2f module increases the depth while ensuring smooth gradients, and the detection accuracy of the model is subsequently improved. However, for medium and small size targets, there are still cases of missed detection, in order to meet the accurate identification of tiny bubbles generated at the subtle defects of defective stainless steel welded pipes, and to increase the detection accuracy without increasing the additional computational cost, this paper proposes a C2f_BoT module for replacing the C2f module in the backbone network, and the structure of the C2f and the C2f_BoT network is shown in Fig. 4.

The BoT(Bottleneck Transformer) modules added after the last layer of Bottleneck in the C2f module, because the BoT module helps to improve the detection accuracy of the model for small and medium-sized targets, in order to maximise the utilisation of computational resources, this paper only replaces the C2f module with the 40×40 and 20×20 scales in the backbone network. The C2f_BoT module can effectively improve the accuracy. Moreover, the generalisation ability of the model can be further improved due to the MHSA attention mechanism in the BoT module. BoT network has a simple and powerful structure and is widely used in visual tasks such as image classification, target detection, semantic segmentation, etc. [22], and its structure is shown in Fig. 5.It replaces the 3×3 convolutional layers in the ResNet structure with the MHSA (Multi-Head Self-Attention), which automatically captures the dependencies between sequences when processing the input sequences and thus better understands the contextual information to improve the model performance. This operation significantly improves the baseline in the target detection task, resulting in lower latency. The C2f_BoT module with the added BoT structure improves the computational efficiency and detection accuracy of the model compared to the original C2f module, which is more suitable for the practical application of the model in industrial inspection.



(a)C2f network structure (b)C2f_BoT network structure

Fig. 4. C2f and C2f_BoT network structure.

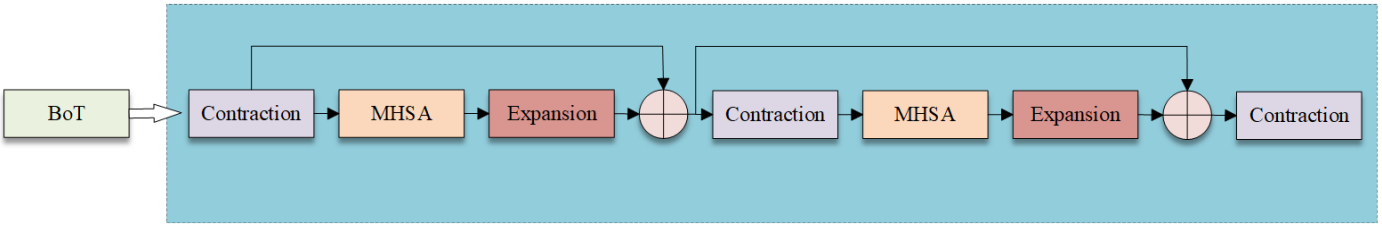


Fig. 5. Bottleneck transform structure.

D. Lightweight Processing

In order to reduce model complexity and reduce computational requirements, this paper uses the GSConv module to replace the original convolutional layer.GSConv is a hybrid of SConv (Standard convolution) and DSConv (Depth-wise separable convolution) [23] combined by the Shuffle convolution. DSConv incorporates deep convolution, which focuses on spatial feature extraction, and point-by-point convolution, which focuses on channel features. This structure performs deep convolution for each channel independently when processing features, and performs channel fusion at the output stage through a 1×1 convolutional layer to reduce the number of computations and parameters. Since DSConv separates each channel at the input image, some of the bubble image features are lost, which is degraded for airtightness detection accuracy. And GSConv can reduce the number of parameters and computation of the model while ensuring the training speed and detection accuracy. Its structure is shown in Fig. 6.

In this paper, the GSConv structure is introduced in the Neck part. In the Neck part, the channel dimension of the feature map extracted by the network is large, in order to maximally retain the circulation of feature information in the spatial and channel dimensions, and to avoid the loss of detail information, the original convolutional layer is replaced with GSConv in this part, which can reduce the model complexity under the premise of guaranteeing a certain model accuracy. If GSConv is used in all parts of the network, it will increase the depth of the network and thus prolong the inference time.

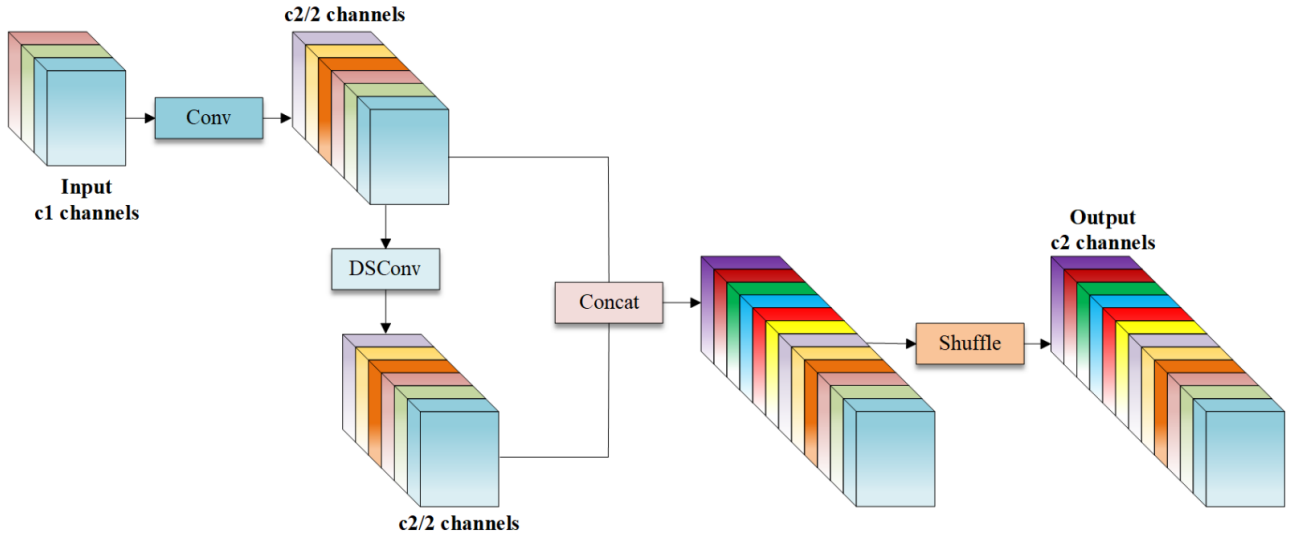


Fig. 6. GSConv structure.

E. C2f-BM

The attention mechanism is a resource allocation scheme that mimics human vision, which effectively allocates computational resources, prioritises critical tasks and alleviates information overload.YOLOv8 often faces the problem of feature loss in stainless steel welded pipe airtightness experiments for detecting small target bubbles, because small bubbles occupy fewer pixels in the feature map, and as the network undergoes many times of downsampling, the feature details of the small bubbles may be overlooked which makes the model performance degraded and generates problems such as leakage detection, which has a significant impact on the leakage detection generated by fine defects in steel pipes. Especially in the complex scene in the water, due to the water surface ripples caused by leakage, the impurity precipitation in the water and the pixel difference between the light and dark areas, the noise information in the background may occupy most of the feature space, so that the model is interfered with and cannot be accurately focused on the target area, resulting in the inability to accurately locate the target, and thus leakage detection.

To solve the above problems, the C2f-BM attention mechanism is designed in this section and embedded in front of the YOLOv8 detection head, the structure of which is shown in Fig. 7.

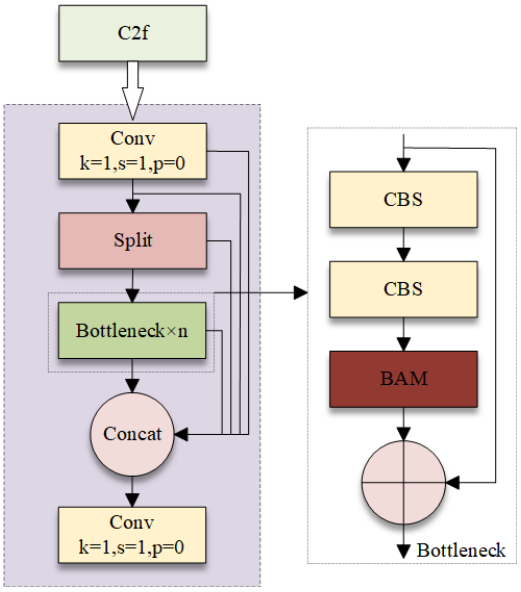


Fig. 7. G2f-BM structure.

The BAM (Bottleneck Attention Module) [24] is a simple and efficient attention mechanism, the structure of which is shown in Fig. 8, which mimics how the human visual system focuses on the critical parts of an image by separating the information from both channel and spatial pathways.

Channel attention focuses on enhancing the identification of feature channels relevant to the target detection task, while spatial attention focuses on the spatial location in the image, helping the model to focus more on the region where the small target is located, thus reducing the loss of feature information of the small target. The BAM combines the results of these two to generate a comprehensive Attention Map. This mechanism effectively improves the model's performance in the detection of small bubble target. This mechanism effectively improves the performance of the model in small bubble target detection, especially in complex detection scenarios in water. Through the spatial attention mechanism, the BAM mechanism can effectively suppress the irrelevant information in the background, reduce the noise interference, and make the model more focused on the feature space of the small target to enhance the robustness of the model. The C2f module itself optimises the feature extraction process, and by embedding the BAM mechanism, the model can focus on important features in the early feature extraction stage to improve the feature quality and model performance in the subsequent stages. Moreover, since C2f improves the gradient flow, the introduction of BAM can further optimise the gradient distribution, which makes the model more stable to be trained when dealing with complex visual tasks. Due to the lightweight nature of the BAM mechanism, it does not significantly increase the extra computational burden during embedding, allowing the whole model to improve performance while maintaining efficient computation.

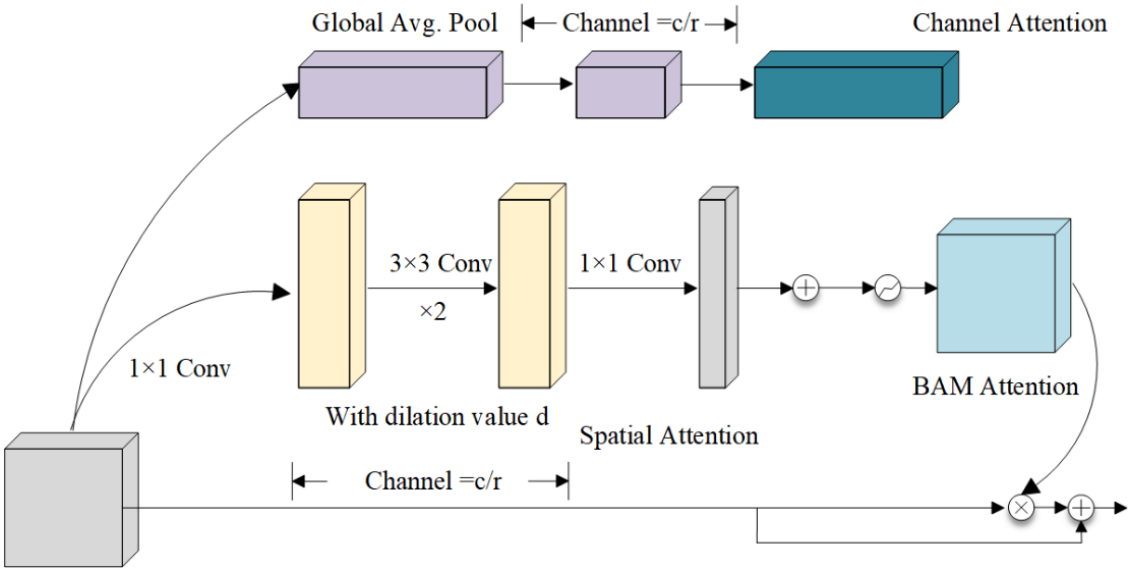


Fig. 8. BAM structure.

BAM calculates Attention Map through two branches: channel and space, and this paper introduces its calculation method from the following three points:

1) .Channel Attention (CA): The expectation maximisation algorithm aims to find maximum likelihood solutions for the hidden variables. In this step, the posterior probability distribution of the hidden variables under the current model parameters is calculated. In fact, it is to calculate the weights and responsibilities of each base for each pixel, the formula is shown in Eq. (1).

$$F_c = AvgPool(F) \tag{1}$$

where the global average pooling of the input feature map $F \in R^{C\times H\times W}$ is performed to obtain the average value of each channel to obtain the vector $F_c \in R^{C\times 1\times 1}$. C is the number of channels, H is the height, and W is the width.

Next, a multilayer perceptron (MLP) containing hidden layers is used to estimate the inter-channel attention from the channel vector F_c ,the formula is shown in Eq. (2).

$$MLP(F_c) = BN(W_1(\mathrm{Re}bLU(W_0F_c+b_0)+b_1)) \tag{2}$$

where, $W_0 \in R^{\frac{C}{r}\times C}$, $b_0 \in R^{\frac{C}{r}}$, $W_1 \in R^{\frac{C}{r}\times C}$, $b_1 \in R^{C}$, the size of the hidden layer is $\frac{C}{r}$, and r is the reduction ratio.

After the MLP, a batch normalisation layer is added to scale the output, the formula is shown in Eq. (3).

$$CA = BN(MLP(F_c)) \tag{3}$$

2) Spatial Attention (SA): First, the feature map F is projected to a smaller size using a 1×1 convolution to integrate and compress the feature map across channel dimensions, the formula is shown in Eq. (4).

$$F_R = Conv_{1\times1}(F) \tag{4}$$

Next, the contextual information is effectively utilised with two 3×3 null convolutions, the formula is shown in Eq. (5).

$$F_D = DilatedConv_{3\times3}(F_R) \tag{5}$$

Finally, the feature map is again reduced to a spatial attention map using 1×1 convolution, the formula is shown in Eq. (6).

$$SA = Conv_{1\times1}(F_D) \tag{6}$$

3) Combination of CA and SA:

$$M(F) = \sigma(M_C(F) + M_S(F)) \tag{7}$$

where $M_C(F)$ is the channel attention mapping, $M_S(F)$ is the spatial attention mapping, and σ is the *Sigmoid* activation function. Finally, the 3D attention map $M(F)$ is multiplied element-by-element with the input feature map, and then added to the original input feature map to obtain the refined feature map $F^{'}$, the formula is shown in Eq. (8).

$$F^{'} = F + F \otimes M(F) \tag{8}$$

where $\otimes$ denotes element-by-element multiplication.

This combination of BAM mechanisms is simple yet effective in adaptively assigning higher weights to small targets by balancing information from both the channel and spatial branches, while facilitating gradient flow.

F. Loss Function Improvement

The original YOLO family of algorithms uses IoU (Intersection over union) to calculate the bounding box regression loss, which refers to the ratio of the intersection and concatenation set of the true and predicted frames, as shown in Eq. (4).

$$IoU = \frac{|A \cap B|}{|A \cup B|} \tag{9}$$

Where A denotes the true frame and B denotes the predicted frame, $A \cap B$ is its intersection, $A \cup B$ is its union set.

When the similarity ratio between the predicted frame and the real frame is higher, it means that the detection is better. However, there are shortcomings, in the detection when the real frame and the predicted frame do not intersect, that is, when $A \cap B = 0$, it is impossible to determine the size of the distance between the predicted frame and the real frame.

Therefore, in order to solve the instability brought by the above situation to the detection, YOLOv8 adopts CIoU (Complete-IoU) [25] instead of IoU. CIoU adds distance and aspect ratio to IoU, and its calculation formula is shown in Eq. (10)- Eq. (12).

$$L_{CIoU} = 1 - IoU + \frac{\rho^2(A,B)}{c^2} + \alpha V \tag{10}$$

$$V = \frac{4}{\pi^2}(\arctan\frac{\omega^A}{\omega^A} - \arctan\frac{\omega^B}{\omega^B}) \tag{11}$$

$$\alpha = \frac{V}{1 - IoU + V} \tag{12}$$

where, $\rho^2(A,B)$ is the Euclidean distance between the centroids of the true and predicted frames, V is a parameter indicating the consistency of the aspect ratio, and α is a parameter used to balance the ratio.

Using CIoU, the minimum value of the distance between the real and predicted frames, the aspect ratio and the distance between the centre points of the bounding boxes can be calculated, which improves the stability of the target box regression. However, CIoU only considers the centre distance of the two bounding boxes, and the matching of the boundaries cannot be accurately assessed, leading to an impact on the detection effect when the target shape changes. The WIoU loss function, on the other hand, introduces a weight function on the basis of IoU, which makes it flexible to adjust the weights between different samples even in more complicated situations. For samples with poorer labelling quality in the dataset, WIoU performs better compared to other boundary loss functions.

In this paper, WIoU (Wise-IoU) is replaced as the loss function, which is proposed on the basis of Focal-EIoU [25].Most of the research on loss function in recent years assumes that the samples in the dataset are of high quality, and is committed to improving the fitting ability of the bounding box loss, while when there are low-quality samples in the dataset, it will jeopardise the model detection performance if the regression of the bounding box on the low-quality samples is improved, while Focal-EIoU is proposed to solve the problem, and its formula is shown in Eq. (13) - Eq. (14).

$$L_{EIoU} = 1 - IoU + \frac{\rho^2(A,B)}{c^2} + \frac{\rho^2(\omega^A,\omega^B)}{c_\omega{}^2} + \frac{\rho^2(h^A,h^B)}{c_h{}^2} \tag{13}$$

$$L_{Focal-EIoU} = IoU^{\gamma} L_{EIoU} \tag{14}$$

where γ is the hyperparameter used to control the curvature of the curve and its focusing mechanism is static. To fully exploit the potential of the non-monotonic focusing mechanism, WIoU [26] uses a dynamic non-monotonic mechanism to assess the quality of the anchor frame, which gives it a better performance when facing the targets with different geometric factors, and the detection performance of the samples from low-quality datasets is improved, and improves the model's generalisation ability. The WIoU loss function expression is shown in Eq. (15) - Eq. (17).

$$L_{WIoU} = \frac{\beta}{\delta\alpha^{\beta-\delta}} R_{WIoU} L_{IoU} \tag{15}$$

$$R_{WIoU} = exp(\frac{(x - x_{gt})^2 + (y - y_{gt})^2}{(\omega_g{}^2 + h_g{}^2)^*})p \tag{16}$$

$$L_{IoU} = 1 - IoU \tag{17}$$

where β describes the outlier of the anchor frame mass, α and δ are hyperparameters. ω_g and h_g are the dimensions of the minimum enclosing frame, (x, y) and (x_{gt}, y_{gt}) are the coordinates of the centre points of the anchor and target frames. Using the WIoU loss function in the improved model, the performance of target detection can be improved by introducing the weighting factor, and for targets with different sizes and shapes, the WIoU can give more reasonable weights to different samples, so that the improved model can better learn the characteristics of bubbles with different sizes and shapes during the training process, and improve the robustness of the stainless steel welded pipe airtightness detection task.

IV. LEIS EXPERIMENTS AND ANALYSIS OF RESULTS

A. Experimental Methodology Flow

The experimental methodology of this paper is specifically divided into the steps of data preparation, model construction, model training, model validation, and result analysis. The flow chart of the experimental method is shown in Fig. 9.

The specific work is as follows:

1) Data preparation. It is necessary to collect the datasets (D1, D2, D3, D4) of stainless steel defective welded pipe leakage detection, perform data enhancement on the D1 and D2 datasets, and divide the datasets into a training set and a validation set.

2) Model construction. Taking YOLOv8n as the base model, the model improvement work in section III of this paper is carried out on it in order to obtain the improved YOLOv8-BGA model.

3) Model training. Train the model on D1 and D2 datasets respectively.

4) Model validation. Use the trained model weights to validate the model on the validation sets of D1, D2, D3, and D4 datasets, and calculate Precision, Recall, and mAP metrics.

5) Result analysis. Compare the detection performance of different models, analyze the generalization ability of models in different environments, and conduct ablation experiments to verify the effectiveness of each improvement module.

B. Experimental Setup

The experiments in this paper were conducted under Windows 11 operating system, using a CPU model i7-13620H with 16G of RAM, a graphics card NVIDIA RTX4060, accelerating the GPU using CUDA11.8 and CUDNN8.8.1, and running under the Pytorch2.0.0 deep learning framework. In this experiment, the official website YOLOv8n model weights are used as the basic network model, batchsize is set to 32 and epoch is 200.an industrial face array camera with model number MV-CS004-10GM is used, and the camera is shown in Fig. 10.

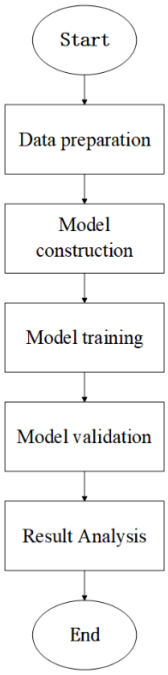


Fig. 9. Experimental methodology.

Fig. 10. Industrial camera.

An airtightness tester model LS11Z-100 was used for inflation and pressurization, as shown in Fig. 11.

Fig. 11. Airtightness tester diagram.

The light source is a strip light source and the schematic diagram of the image acquisition device is shown in Fig. 12.

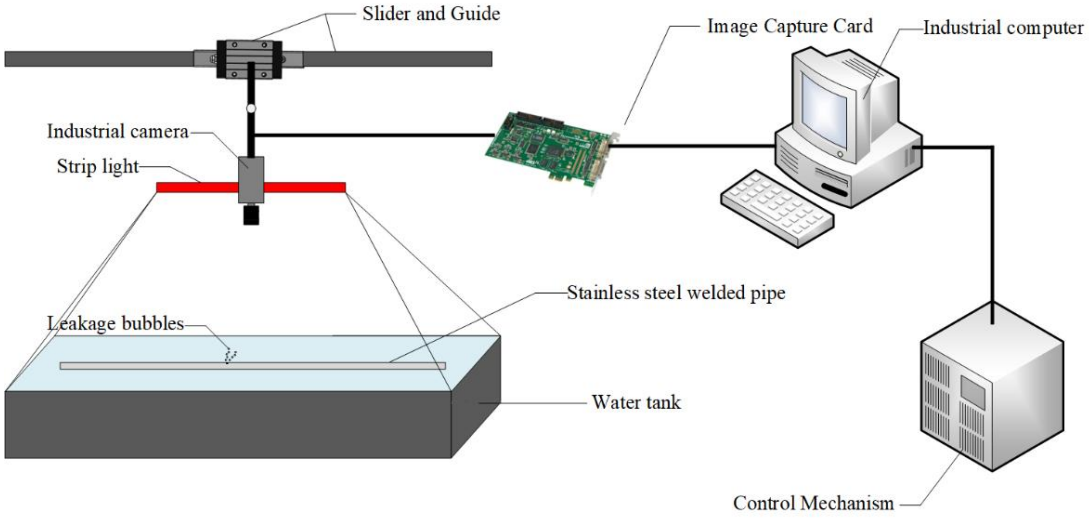
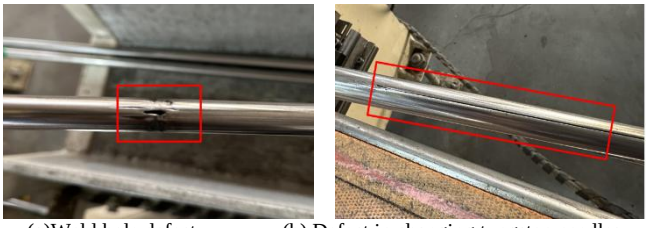


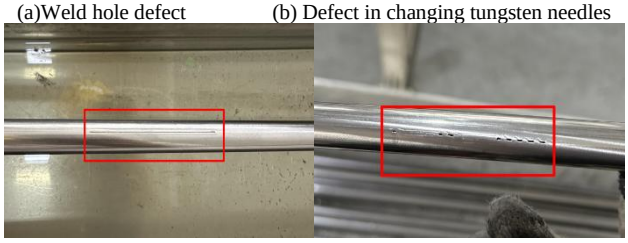
Fig. 12. Image acquisition device.

C. Experimental Data Set

The dataset in this paper was acquired in an image acquisition platform built in a real industrial inspection environment. The stainless steel defective welded pipe used for the experiment is obtained from the processing workshop, and several common defective welded pipes are shown in Fig. 13.



(a)Weld hole defect (b) Defect in changing tungsten needles



(c) Soldering defect (d) Skipped weld defects

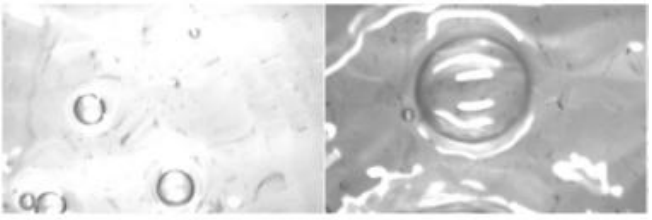
Fig. 13. Stainless steel welded pipe common defects diagram.

In order to evaluate the generalisation ability of the improved model, two datasets, D1 and D2, were acquired in two different lighting environments (bright and dimmer), and secondly, two noisy images were acquired in these two lighting environments as D3 and D4 datasets. All the dataset images were labelled and their basic characteristics are detailed in Table I.

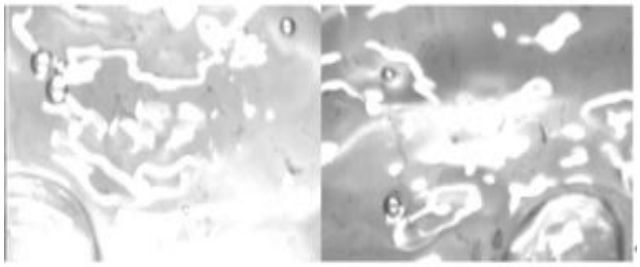
TABLE I BASIC INFORMATION ON THE D1-D4 DATA SETS

Dataset	Feature	train set	val set
D1	Sufficient light	4640	1160
D2	dusky	5080	1270
D3	Sufficient light, noise	0	945
D4	Dusky, noise	0	975

Due to the defects of the original data, Mosaic data enhancement is performed on D1 and D2 datasets, and the images of D1 dataset are expanded to 5800, D2 dataset is expanded to 6350, and D3 and D4 datasets are not processed. The D1 and D2 datasets are divided into training set and validation set with the ratio of 8:2. Some images of D1-D4 datasets are shown in Fig. 14.



D1 D2



D3 D4

Fig. 14. Partial images of the dataset.

D. Assessment of Indicators

In the field of deep learning target detection, the main metrics for model performance evaluation are P (Precision), R (Recall) and mAP (mean AP), and the model complexity can be evaluated by GFLOPS (Giga Floating Point Operations) and Parameters [27]. FPS were used to evaluate the performance of the model. In this paper, the model evaluation is also based on these criteria.p, R and mAP are calculated as shown in Eq. (18)-Eq. (21):

$$P = \frac{TP}{TP + FP} \tag{18}$$

$$R = \frac{TP}{TP + FN} \tag{19}$$

$$AP = P(r)dr \tag{20}$$

$$mAP = \frac{1}{N} AP \tag{21}$$

In Eq. (13) - Eq. (16), TP is the number of positive samples correctly detected; FP is the number of positive samples incorrectly detected; FN is the number of negative samples incorrectly detected; AP is the average precision; and N is the number of all predictions.

E. Error Analysis and Discussion

In the process of detecting leakage bubbles in stainless steel welded pipes using the YOLOv8-BGA model, several types of errors were observed. Small bubbles, particularly those generated by subtle defects, are more likely to be missed due to their low pixel occupancy in the feature map. The repeated downsampling in the network can lead to the loss of these fine details. Additionally, bubbles that appear in regions with high background noise or similar textures are prone to incorrect detection. For example, in the presence of water impurities or ripples, the model may confuse these with actual bubbles. These factors collectively contribute to the challenges faced during the detection process.

The impact of noise and ripples on detection accuracy cannot be overlooked. The movement of bubbles can generate ripples on the water surface, which introduce additional noise into the image. These ripples can interfere with the model's ability to accurately locate and identify bubbles, especially in complex underwater scenes. Variations in lighting conditions can cause significant reflection and refraction effects on the water surface, leading to false positives or false negatives as the model may misinterpret the light patterns as bubbles. Overlapping bubbles and those with unclear or irregular boundaries also pose challenges, as the model may struggle to distinguish individual bubbles or accurately delineate bubble contours. These issues highlight the need for further improvements to enhance the model's robustness and accuracy.

To address these challenges, several strategies were implemented. Mosaic data augmentation was employed to increase the diversity of the training dataset, helping the model learn to detect small targets more effectively. The C2f-BM attention mechanism was introduced to allow the model to focus more on critical regions, reducing the impact of noise and improving the detection accuracy for small bubbles. Additionally, the WIoU loss function was utilized to enhance the model's fitting ability, especially for low-quality samples with small bubbles, thereby improving the overall detection performance. Through these improvements, the YOLOv8-BGA model demonstrated enhanced robustness and accuracy in detecting leakage bubbles under various conditions. Future work will continue to explore advanced techniques and additional data collection under varied conditions to further improve the model's performance.

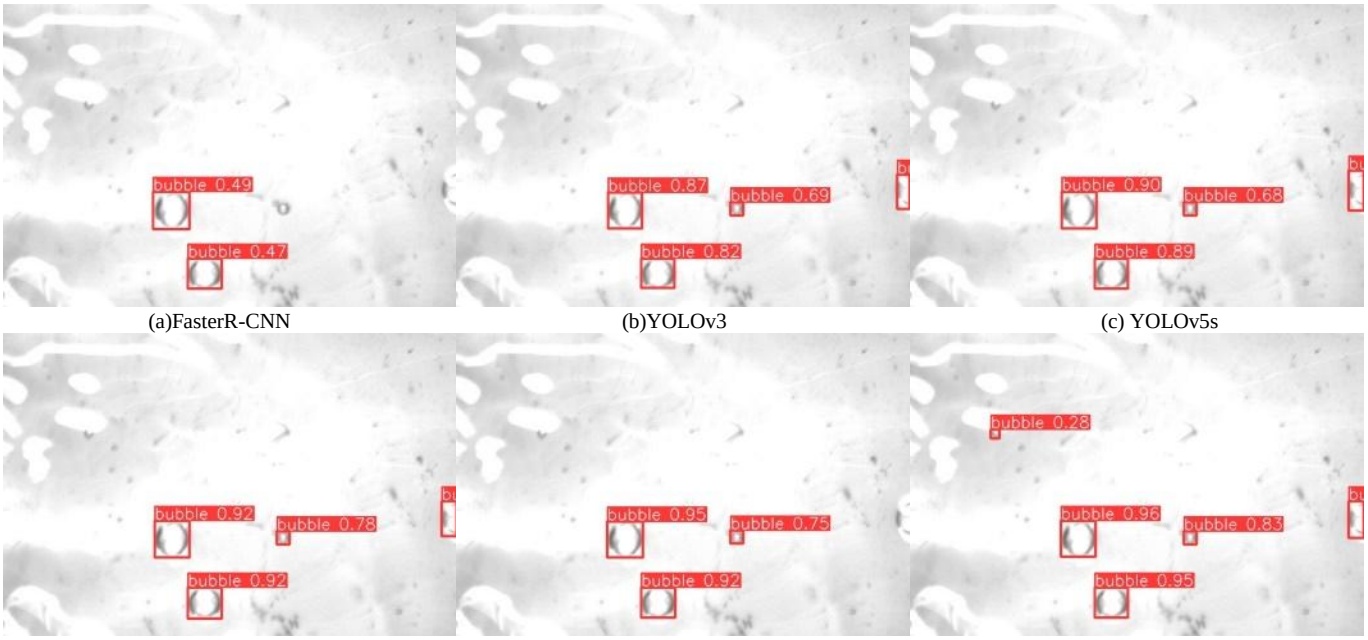
F. Experimental Results and Analysis

1) Comparison experiment: In order to verify the performance of the improved model in this paper, a comparison experiment is first taken. Several common target detection algorithms are selected and trained on D1 dataset under the same experimental environment and experimental configuration, and validated. The experimental results are detailed in Table II.

TABLE II COMPARATIVE EXPERIMENTAL RESULTS OF SOME MAINSTREAM ALGORITHMS

Model	P/%	R/%	mAP/%t	Parameters/M	FLOPS/G	FPS (F/S)
Faster R-CNN	60.1	63.7	64.6	14.3	19.8	25
YOLOv3	84.6	85.4	87.3	54	125.6	45
YOLOv5s	90.1	89.8	91.6	7.0	15.8	65
YOLOv8n	90.6	90.8	92.4	3.0	8.1	85
YOLOv8s	89.9	90.4	91.5	11.1	28.4	50
YOLOv8-BGA	95.4	96.3	97.7	2.7	7.5	89

Table II shows that the improved algorithm in this paper has excellent training effect on the D1 dataset, and has certain improvement compared with other algorithms. Compared with Faster R-CNN, it has the biggest improvement, with an average accuracy improvement of 34.1%; compared with YOLOv3, the average accuracy improvement is 11.4%; compared with YOLOv5s and YOLOv8s, it has an improvement of 7.1% and 7.2%, respectively; this paper improves the improved model by using YOLOv8n as the baseline, and the improved model has an average accuracy improvement of 5.3% compared with YOLOv8n.The FPS value of YOLOv8-BGA is 89, which is significantly higher than that of other algorithms, indicating that it not only improves the detection accuracy, but also optimizes the inference speed of the model, which can make the detection task achieve a good balance between real-time and accuracy. In addition, while the detection accuracy is improved, the number of parameters and inference time of the model are lower than the other five algorithms. It can be seen that YOLOv8-BGA has higher detection accuracy in the dataset of this paper, and at the same time, the model complexity is lower, which can better meet the actual industrial detection needs of stainless steel defective welded pipe bubbles. The comparison of the detection results of YOLOv8-BGA and the other algorithms is shown in Fig. 15.



(a)FasterR-CNN (b)YOLOv3 (c) YOLOv5s

(d) YOLOv8n (e) YOLOv8s (f) YOLOv8-BGA

Fig. 15. Comparison of detection results between YOLOv8-BGA and other algorithms.

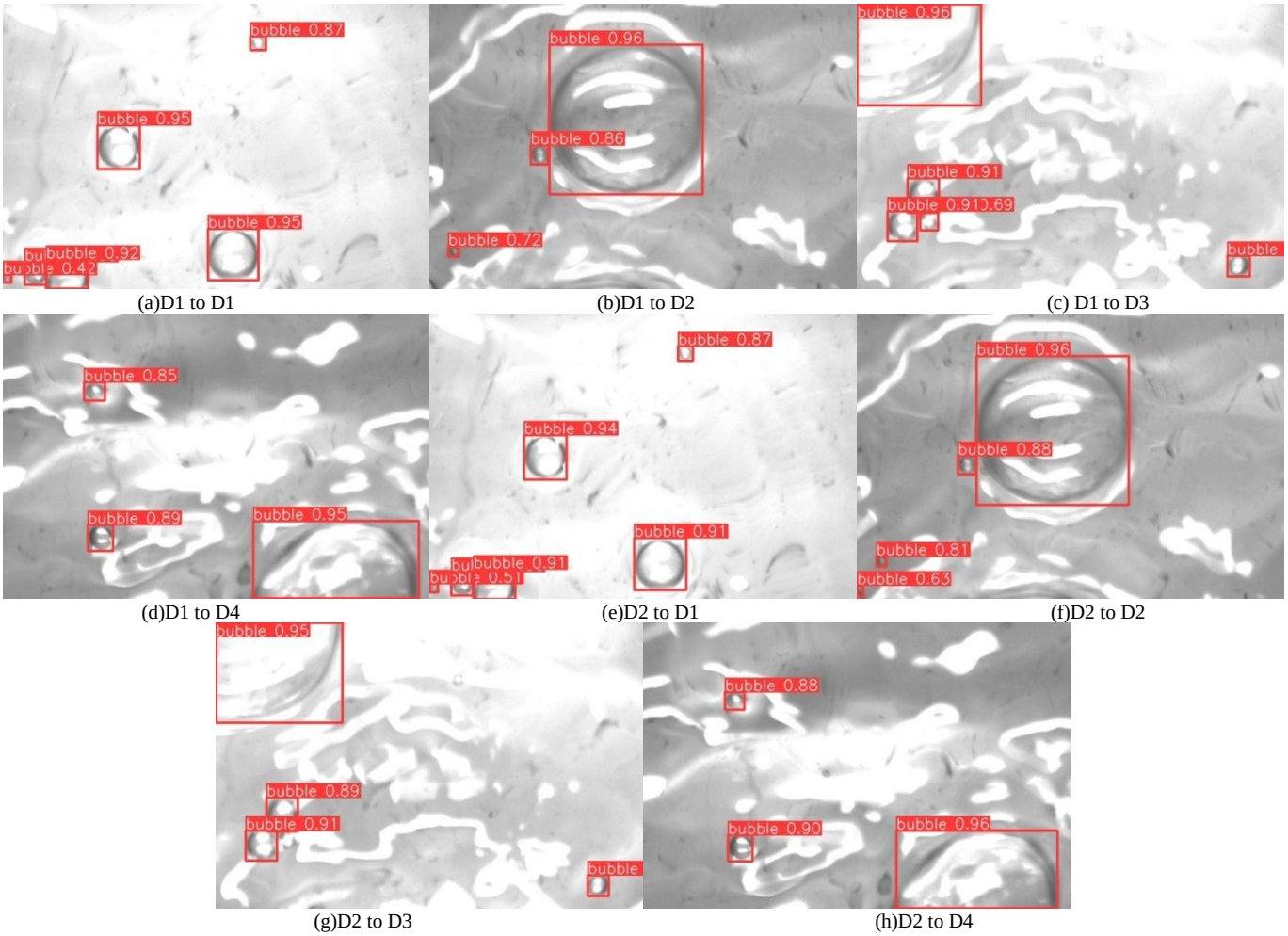
2) Verification of generalisation capabilities: In the actual detection process, the image quality is affected by a variety of factors such as illumination, water reflection and ripples, so it is crucial to evaluate the generalisation ability of the model. In this section of experiments, the improved algorithm is trained on D1 and D2 datasets under the same configuration, and the corresponding validation sets are validated with the respective training results. Precision and recall are chosen to evaluate the model performance, and the experimental results are detailed in Table 3.In this experiment, the influence of lighting conditions on the leakage bubble detection results of stainless steel welded pipe based on YOLOv8-BGA model is emphatically analyzed. In this section, two lighting environments is set up, bright and dark, and the results show that the lighting conditions have a significant impact on the model detection performance. The results show that the lighting conditions have a certain influence on the model detection performance.

TABLE III COMPARISON OF TEST RESULTS

mission	P/%	R/%
D1 Validated D1	95.4	96.3
D2 Validated D2	93.9	94.3

In low-illumination conditions, the overall brightness of the image is reduced, which directly leads to a decrease in image contrast. According to the fundamental principles of image processing, contrast is a key factor in distinguishing targets from the background. When contrast is reduced, the differences between small bubbles and the background become less distinct, especially for small bubbles generated by minor defects, which occupy fewer pixels in the feature map. In low-contrast images, the features of these small bubbles become even more difficult to extract. For example, in our experimental dataset, when the illumination is dim, the edge details of small bubbles are blurred, making it challenging for the model to accurately identify their contours. This results in larger deviations between the predicted and ground-truth bounding boxes when calculating the loss function, thereby reducing the detection accuracy of the model. Specifically, when validating on the dataset D2 under low-illumination conditions, the model's Precision is lower compared to that on the dataset D1 under bright illumination conditions. For instance, the Precision is 95.4% when D1 is validated on D1, while it is 93.9% when D2 is validated on D2.In bright illumination conditions, although the overall brightness of the image is sufficient, strong reflections become a significant factor affecting detection results. When light strikes the surface of the water and bubbles, reflection and refraction occur. According to the principles of optics, the angle of incidence is equal to the angle of reflection. When light is incident at a large angle, the intensity of the reflected light increases. Under bright illumination, due to the undulations of the water surface and the movement of bubbles, the direction of the reflected light constantly changes, causing the bubble boundaries to appear blurred in the image. This blurred boundary interferes with the model's accurate judgment of bubbles because the model relies on clear boundary features for localization and identification. For example, in the experiment, we observed that under bright illumination with undulating water surfaces, the model is prone to misjudging reflective areas as bubbles or inaccurately locating the boundaries of bubbles, thereby affecting Recall and Precision.

Then, in order to verify the generalisation ability of the improved algorithm under different environmental conditions, the D2, D3, and D4 datasets were validated using the training results of the D1 training set, and the D1, D3, and D4 were validated using the training results of the D2 training set, respectively, and the results are shown in Fig. 16.



(a)D1 to D1 (b)D1 to D2 (c) D1 to D3

(d)D1 to D4 (e)D2 to D1 (f)D2 to D2

(g)D2 to D3 (h)D2 to D4

Fig. 16. Comparison of some test results.

In assessing the performance of the model in various environments, precision and recall are also used as evaluation metrics, as detailed in Table IV.

TABLE IV RESULTS OF DIFFERENT ENVIRONMENTS

Mission	P/%	R/%
D1 Validated D2	88.7	85.6
D1 Validated D3	92.4	93.6
D1 Validated D4	85.9	86.1
D2 Validated D1	86.9	83.2
D2 Validated D3	83.4	82.9
D2 Validated D4	90.8	91.5

Comparing Table IV, it can be found that the validation is carried out under datasets with different environments, and although the precision and recall of the improved algorithm have decreased compared with those under the corresponding datasets, the model as a whole still maintains a good performance. The experimental results illustrate that the improved YOLOv8-BGA algorithm has good generalisation ability under different detection environments, and can effectively detect air bubbles generated by stainless steel defective welded pipes in real time under a variety of detection environments. Using the model trained under the D1 dataset to validate the D3 dataset is 9% more accurate compared to the model trained under the D2 dataset to validate the D3 dataset; similarly, using D1 to validate D4 is 4.9% less accurate compared to D2 to validate D4, which indicates that the performance of the improved model trained under the D1 dataset is higher than the performance of the model trained under the D2 dataset. Because the D1 dataset and the D3 dataset, the D2 dataset and the D4 dataset have the same kind of lighting conditions, which indicates that the lighting conditions have a greater impact on the training results of this model, and the image quality of the D1 dataset is higher than that of the D2, which may affect the performance of the model generalisation ability.

3) Comparative experiments with multiple datasets: To further validate the generalization and robustness of the improved model across diverse fields and industries, this section devises experiments to assess the performance of YOLOv8n and YOLOv8-BGA on publicly available datasets, namely the Bubble Image Database and UF-120.The Bubble Image Database, disseminated by researchers from the University of Queensland, encompasses 5,184 original images and 25,920 augmented images. It comprises five bubble categories: “Fine bubbles fully covering the viewing window”,

"fine bubbles partially covering the viewing window", "Co-existence of coarse and fine bubbles", "Only coarse bubbles", and "No bubbles". The UF-120 dataset consists of 120 high-quality underwater bubble images, representing a wide array of underwater scenarios and faithfully mirroring the complexity inherent in underwater environments. The experimental outcomes are presented in Table V and Table VI.

TABLE V EXPERIMENTAL RESULTS OF BUBBLE IMAGE DATABASE

Model	mAP/%	Parameters/M	FLOPS/G	FPS(F/S)
YOLOv8n	82.1	3.1	8.2	70
YOLOv8-BGA	82.7	2.8	7.6	72

TABLE VI EXPERIMENTAL RESULTS OF UF-120

Model	mAP/%	Parameters/M	FLOPS/G	FPS(F/S)
YOLOv8n	89.1	3.0	8.2	77
YOLOv8-BGA	89.8	2.7	7.6	75

Evident from the findings presented in Table V, the YOLOv8-BGA model exhibits an increment of 0.6 in the mAP. Concurrently, the model experiences a reduction of 9.7% in the Parameters and 7.3% in the FLOPS, with no discernible alteration in the detection speed. Inspection of the results in Table 6 reveals that the YOLOv8-BGA model registers a mAP increase of 0.7, accompanied by decreases of 10% in Parameters and 7.3% in FLOPS. The detection speed remains commendably stable, showing no significant fluctuations. Experimental outcomes on publicly accessible datasets, including the Bubble Image Database and UF-120, incontrovertibly demonstrate that, within the domain of bubble detection, the YOLOv8-BGA model surpasses the YOLOv8n model with respect to accuracy, model complexity, and computational efficiency. Overall, the YOLOv8-BGA model demonstrates good utility value and robustness in practical industrial applications.

4) Ablation experiment: In order to test the effectiveness and model performance of adding BoT module, replacing GSConv module, adding C2f-BM attention mechanism and improving loss function in the improved algorithm on the gas tightness detection of stainless steel defective welded pipe, ablation experiments are carried out on YOLOv8-BGA algorithm under D1 dataset, and the results of the ablation experiments are shown in Table VII in detail.

Analysing the experimental results in Table VII, the YOLOv8n model is used as the baseline, and the improved modules are added individually in sequence for comparison.YOLOv8-BoT replaces the 40×40 and 20×20 C2f modules in the Backbone section with the C2f_BoT module, which improves the average accuracy by 0.9% without any change in the number of Parameters and FLOPS, which is due to the fact that the BoT module is able to automatically identify and capture the dependencies in the input sequence through its multi-head self-attention mechanism, which helps the model to extract features more efficiently and improves the detection accuracy, the introduction of the BoT module did not significantly increase the model complexity, so the FPS was not significantly reduced; YOLOv8-GSConv is to replace the original convolutional layer in the Neck part with the GSConv module, and the number of parameters is reduced by about 0.2M, and the FLOPS is reduced by 0.4G.The GSConv module reduces the amount of computation through deeply separable convolution, which improves the lightness of the model and the speed of inference, so the FPS is improved, the model achieves a lightweight effect while maintaining the detection accuracy; YOLOv8-BM is to add the C2f-BM attention mechanism in front of each detection layer, the number of parameters is reduced by about 0.1M, the FLOPS is reduced by 0.2G, and the average accuracy is improved by 4.4%, which indicates that the expectation maximisation algorithm reduces the computational volume of the model while the detection accuracy is improved dramatically, the attention mechanism performs well on this dataset, improving performance along with inference speed and FPS values; YOLOv8-WIoU introduces the WIoU loss function, and the average accuracy of the model is improved by 1.9%, and the WIoU loss function improvesthe degree of model fitting, which in turn improves the accuracy of model recognition. The introduction of the WIoU loss function, which mainly optimizes the training process, has a small impact on the inference speed, so the FPS value is unchanged; YOLOv8-BGA is the final model after adding all of the above modules, and the number of references is lowered by about 0.3M, FLOPS is reduced by 0.6G, and the average accuracy is improved by 5.3%.After combining all the improvements, the model is optimized in all aspects, and the inference speed is further optimized while maintaining high accuracy, and the final FPS value of YOLOv8-BGA reaches 89.

TABLE VII RESULTS OF ABLATION EXPERIMENTS

Model	BoT	GSConv	C2f-BM	WIoU	mAP/%	Parameters/M	FLOPS/G	FPS(F/S)
YOLOv8n					92.4	3.0	8.1	85
YOLOv8-BoT	√				93.3	3.0	8.1	84
YOLOv8-GSConv		√			92.1	2.8	7.7	87
YOLOv8-BM			√		96.8	2.9	7.9	86
YOLOv8-WIoU				√	94.3	3.0	8.1	85
YOLOv8-BGA	√	√	√	√	97.7	2.7	7.5	89

The experimental results verify that the YOLOv8-BGA model obtained after the improvement of this paper has been improved compared with the original model in terms of reasoning speed, detection accuracy, model complexity, etc., and is able to better complete the task of stainless steel defective welded pipe airtightness detection compared with other algorithms proposed above.

V. CONCLUSION

1) The improved model is better than the original YOLOv8n model in terms of inference speed, detection accuracy, and model complexity, and can achieve effective identification of leakage bubbles.

2) With the introduction of the C2f-BoT module and the WIoU loss function, the improved model possesses excellent detection capability for tiny bubbles and reduces leakage detection.

3) The improved model has a strong generalisation ability, and can also have a good detection ability in poorer airtightness detection environments.

4) Due to the limitations of the experimental conditions, the accurate linkage between real-time detection, positioning and alarm at the leakage of stainless steel welded pipe needs to be further investigated in order to improve the popularisation of the improved model in the field of industrial airtightness detection.The model algorithm is a useful exploration and reference for other different products airtightness detection.

REFERENCES

[1] Li, L., Yu, T., Xia, J., Gao, Y., Han, B., Gao, Z.: Failure analysis of L415/316L composite pipe welded joint.Engineering Failure Analysis, 158, 107981.(2024)

[2] Jiao, C., Zhang, R., Hou, H., Zhao, Z., Tang, C.and Yun, H.: Research on ultrasonic non-destructive evaluation algorithm for ultimate tensile strength of stainless-steel welded pipe welds based on width learning.Nondestructive Testing and Evaluation, pp.1-15.(2024)

[3] Heim, D.and Miszczuk, A.: Modelling building infiltration using the airflow network model approach calibrated by air-tightness test results and leak detection.Building Services Engineering Research and Technology, 41(6), pp.681-693.(2020)

[4] F.Gao, J.Lin, Y.Ge, S.Lu and Y.Zhang.: A Mechanism and Method of Leak Detection for Pressure Vessel: Whether, when, and how.IEEE Transactions on Instrumentation and Measurement 69(9): 6004-6015.(2020)

[5] Luo, Q., Fang, X., Liu, L., Yang, C.and Sun, Y.: Automated visual defect detection for flat steel surface: A survey.IEEE Transactions on Instrumentation and Measurement, 69(3), pp.626-644.(2020)

[6] Ren, Z., Fang, F., Yan, N.and Wu, Y.: State of the art in defect detection based on machine vision.International Journal of Precision Engineering and Manufacturing Green Technology, 9(2), pp.661-691.(2021)

[7] Qaddoori, A.S., Saud, J.H., Hamad, F.A.: A classifier design for micro bubble generators based on deep learning technique.Materials Today: Proceedings, 2023, 80: 2684-2696.(2023)

[8] Wen, J., Sun, Q., Sun, Z., Gu, H.: An improved image processing technique for determination of volume and surface area of rising bubble.International Journal of Multiphase Flow, 104, 294-306.(2018)

[9] Akdemir, B., Öztürk, S.: Glass surface defects detection with wavelet transforms.International Journal of Materials, Mechanics and Manufacturing ,3(3): 170-173.(2015)

[10] Xu, J., Ren, H., Cai, S., Zhang, X.: An improved faster R-CNN algorithm for assisted detection of lung nodules.Computers In Biology and Medicine, 153: 106470.(2023)

[11] Wang, W., Xu, X., Yang, H.: Intelligent Detection of Tunnel Leakage Based on Improved Mask R-CNN.Symmetry, 2024, 16(6): 709.(2024)

[12] Krysko, N.V., Shchipakov, N.A., Kozlov, D.M., Kusyy, A.G., & Skrynnikov, S.V.: Classification and Sizing of Surface Defects in Pipelines Based on the Results of Combined Diagnostics by Ultrasonic, Eddy Current, and Visual Inspection Methods of Nondestructive Testing.Russian Journal of Nondestructive Testing, 59(12): 1315-1323.(2023)

[13] Zhao Q, Zheng C, Ma W.: An Improved Crucible Spatial Bubble Detection Based on YOLOv5 Fusion Target Tracking.Sensors, 22(17): 6356.(2022)

[14] Redmon, J., Divvala, S., Girshick, R.and Farhadi, A.: You only look once: Unified, real-time object detection.Proceedings of the IEEE conference on computer vision and pattern recognition.(2016)

[15] Li, M., Zhang, J., Liu, H., Yuan, Y., Li, J., Zhao, L.: A lightweight method for apple-on-tree detection based on improved YOLOv5.Signal, Image and Video Processing, 1-15.(2024)

[16] Sun, F., He, N., Wang, X., Liu, H., Zou, Y.: YOLOv7-P: a lighter and more effective UAV aerial photography object detection algorithm.Signal, Image and Video Processing, 1-9.(2024)

[17] Zhao, S., Tao, R., Jia, F.: DML-YOLOv8-SAR image object detection algorithm.Signal, Image and Video Processing, 1-13.(2024)

[18] Zhang, Y., Zhang, H., Huang, Q., Han, Y., Zhao, M.: DsP-YOLO: An anchor-free network with DsPAN for small object detection of multiscale defects.Expert Systems with Applications, 241: 122669.(2024)

[19] Yan, J., Wang, H., Yan, M., Diao, W., Sun, X., Li, H.: IoU-adaptive deformable R-CNN: Make full use of IoU for multi-class object detection in remote sensing imagery.Remote Sensing, 11(3): 286.(2019)

[20] Kang, S.H., Palakonda, V., Kim, I.M., Kang, J.M.and Yun, S.: Enhanced Non-Maximum Suppression for the Detection of Steel Surface Defects.Mathematics, 11(18): 3898.(2023)

[21] Zou, Y.and Fan, Y.: An Infrared Image Defect Detection Method for Steel Based on Regularized YOLO.Sensors, 24(5): 1674.(2024)

[22] Nakai, K., Chen, Y.W., Han, X.H.: Enhanced deep bottleneck transformer model for skin lesion classification.Biomedical Signal Processing and Control, 78: 103997.(2022)

[23] Huang, D., Tu, Y., Zhang, Z., Ye, Z.: A Lightweight Vehicle Detection Method Fusing GSConv and coordinate attention mechanism.Sensors, 24(8): 2394.(2024)

[24] Chen, Z., Tian, S., Yu, L., Zhang, L., Zhang, X.: An Object Detection Network Based on YOLOv4 and Improved Spatial Attention Mechanism.Journal of Intelligent & Fuzzy Systems, 42(3), 2359-2368 (2022)

[25] Aswal, D., Shukla, P., Nandi, G.C.: Designing effective power law-based loss function for faster and better bounding box regression.Machine Vision and Applications, 32(4): 87.(2021)

[26] Yang, X., Liu, C., Han, J.: Reparameterized underwater object detection network improved by cone-rod cell module and WIOU loss.Complex & Intelligent Systems, 1-16.(2024)

[27] Li, S., Zhang, X., Shan, R.: Enhanced YOLOv5 for Efficient Marine Debris Detection.Engineering Letters, 32(8).(2024)