aerospace

MDPI

Article

MobGSim-YOLO: Mobile Device Terminal-Based Crack Hole Detection Model for Aero-Engine Blades

Xinyao Hou *, Hao Zeng, Lu Jia, Jingbo Peng and Weixuan Wang

Aviation Engineering School, Air Force Engineering University, Xi'an 710038, China;
18403435829@163.com (H.Z.); 1217801420@qq.com (L.J.); pjb1209@126.com (J.P.); 18149069921@163.com (W.W.)
* Correspondence: 19192411230@163.com

Abstract: Hole detection is an important means of crack detection for aero-engine blades, and the current technology still mainly relies on manual operation, which may cause safety hazards for visual reasons. To address this problem, this paper proposes a deep learning-based, aero-engine blade crack detection model. First, the K-means++ algorithm is used to recalculate the anchor points, which reduces the influence of the anchor frame on the accuracy; second, the backbone network of YOLOv5s is replaced with Mobilenetv3 for a lightweight design; then, the slim-neck module is embedded into the neck part, and the activation function is replaced with Hard Sigmoid for redesign, which improves the accuracy and the convergence speed. Finally, in order to improve the learning ability for small targets, the SimAM attention mechanism is embedded in the head. A large number of ablation tests are conducted in real engine blade data, and the results show that the average precision of the improved model is 93.1%, which is 29.3% higher; the number of parameters of the model is 12.58 MB, which is 52.96% less, and the Frames Per Second (FPS) can be up to 95. The proposed algorithm meets the practical needs and is suitable for hole detection.

Keywords: aero-engine; non-destructive testing; YOLOv5; lightweighting

check for updates

Citation: Hou, X.; Zeng, H.; Jia, L.; Peng, J.; Wang, W. MobGSim-YOLO: Mobile Device Terminal-Based Crack Hole Detection Model for Aero-Engine Blades. *Aerospace* **2024**, *11*, 676. https://doi.org/10.3390/aerospace11080676

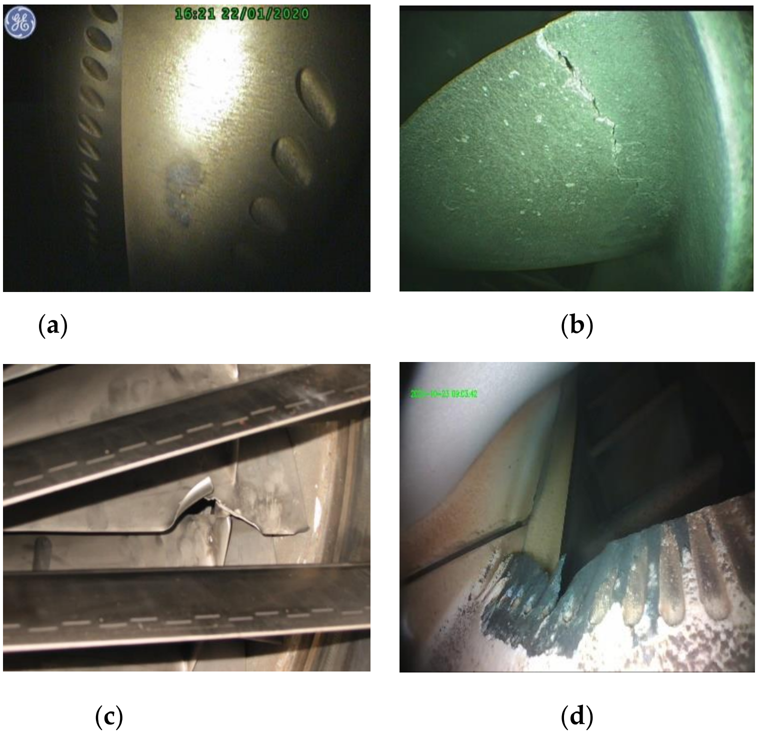
Academic Editor: Cheng-Wei Fei

Received: 12 June 2024
Revised: 4 July 2024
Accepted: 5 July 2024
Published: 16 August 2024

1. Introduction

Since the heart of aircraft aviation engines usually works within high-temperature, high-pressure, high-load, high-speed, strongly corrosive, and harsh environments, the core components inside the engine's applicability and reliability will decline, affecting the aircraft's long-term capability [1]. High-speed rotors, including compressor and turbine blades, often undergo deformation, bruising, tearing, cracking, discoloration under high temperatures, and other structural damage. These are shown in Figure 1. The sources of damage mainly include centrifugal inertia force generated by high-speed rotation, aerodynamic force generated by flowing gas, fatigue damage due to long operation time, and impact of external objects [2]. Therefore, timely and accurate detection and resolution of engine blade structural losses during the maintenance period not only extends the service life of the aircraft but also ensures personnel safety [3].

At present, non-destructive testing has been widely applied to detect structural damage to engine blades [4]. Wang et al. [5] used ultrasonic technology to detect small cracks in blades. They first fabricated blades with different degrees of damage on the vibration experimental bench and then utilized non-linear ultrasonic testing to detect non-linear coefficients of different faulty blades. Through a large number of experiments and statistical analyses, they derived the empirical non-linear coefficients and the equivalent crack size formula, which can be used for blade crack evaluation. Based on the principle of electromagnetic induction, eddy current detection can be used to detect faulty blades. Xie et al. [6] designed a new type of flexible eddy current array sensor that generates eddy currents through the excitation and sensor coils. When the eddy currents pass through the damaged part of the blade, the eddy currents change significantly to derive the type, location, and size of the damage to the blade. Yang et al. [7] also designed an eddy current automatic detection system with six degrees of freedom, and the experiments show that the detection sensitivity of the automatic detection system is very high. Liu et al. [8] also utilized digital radiography for non-destructive testing of gas turbine blades, with remarkable results. Karatay et al. [9] investigated Fluorescein isothiocyanate-conjugated Escherichia coli as a penetrant that can be used to detect leaf cracks.



(**a**) (**b**)

(**c**) (**d**)

Figure 1. (**a**) Turbine blade corrosion, (**b**) Cracked compressor blade, (**c**) Torn compressor blade, (**d**) High-temperature ablation of the turbine blade.

Due to the many types of engine components and complex structure, it is difficult and costly to dismantle the engine. Hole probe inspection enters the engine interior with a camera through a small hole in the engine to inspect the blades in real-time, which can avoid disassembling the engine. By changing the probe insertion depth, rotating the probe direction and engine rotor angle, maintenance personnel can collect video and image information of different levels of blades and different angles of the rotor and then analyze and judge the state of the blades [10]. However, these methods still rely on visual inspection. There is a subjective factor. Different mechanics detecting the same damaged blade may have different conclusions, with a certain risk. The engine's internal light is dark, the contrast is low, it is easy to hide the hidden faults, the number of blade stages inside the engine is large, and the overhauling workload is large, which is a very high requirement for the maintenance personnel [11].

In recent years, artificial intelligence technology has continued to develop and is widely used in various industries. Target detection technology has a strong feature learning ability [12], which can be combined with hole detection technology to help maintenance personnel identify and process image information, make more accurate and rapid judgment decisions, and improve maintenance ability. The current popular target detection algorithms are ResNet [13], R-CNN series [14], YOLO series [15], FCN series [16] and Mask-RCNN [17], etc.

More and more scholars have now applied target detection techniques to engine borehole inspection. Anurag et al. [18] designed the U-Net architecture to detect defects on high-pressure compressor blades, and the detection precision and recall rate exceeded 90%. Still, the detection sensitivity was low for small-size defects. Li et al. [19] combined high-resolution engine blade images into a deep convolutional neural network (DCNN) and then proposed a coarse classifier over most of the background parts. Finally, the defects were detected by a fine detector module, which showed a better detection performance. Zhang et al. [20] applied YOLOv3 to the task of detecting damage on aero-engine blades and achieved a balance between detection accuracy and detection speed. Li et al. [21] proposed an improved intelligent detection model for YOLOv4, which fuses shallow and deep features, improves the PANet structure, and employs focusing loss, and the results show that the improved model achieves an average accuracy of 90.1% and an FPS of 24.82, which falls short of real-time detection. Li et al. [22] introduced deformable convolution and depth-separable convolution on the basis of YOLOv5s and used k-means clustering to optimize the anchor frame, and the results showed that the detection precision could reach 93.3% and the recall rate could reach 76.2%, but the parameter count of the model was 7,928,117, the weight file size was 15.3 MB, and the average detection time was 28.8 ms. The model has high precision, but the model parameters are too large, the detection speed is slow, and further improvement is needed. Cai et al. [23] reconstructed the structure of YOLOv5, changed the backbone network to FasterNet, and introduced the depth separable convolution and GSConv in the neck. The results show that the improved model reduces the number of parameters by 52.5% compared to the YOLOv5 model, the FPS reaches 61, and the average accuracy value reaches 89.6%, and there is still much room for improvement in the accuracy and speed of detection of this model. Shang et al. [3] constructed an enhanced shape Mask R-CNN network with three functions: damage pattern separation, damage localization, and damage region segmentation. The model pays more attention to more attention to the texture information model; although the detection accuracy is higher, it belongs to the two-stage, which may bring about the problem of high computational cost and slow detection speed, and it still can’t be better luck to use in the hole probing detection. Li et al. [24] embedded the CBAM attention mechanism module into YOLOv7 and utilized Alpha-GIOU as the coordinate damage function. The average accuracy of the improved model was 96.1%, but the number of parameters was 36.52 MB, and the FPS reached 85.92; although the accuracy of the model was higher, the number of parameters of the model was larger, and it brought high requirements to the mobile device. In summary, although many scholars have used the target detection technology for engine borehole exploration equipment, they are unable to balance the relationship between test accuracy, test speed, and model size; there is still a lot of room for development. Therefore, this paper combines the current mainstream improvement module reorganization YOLOv5s model to try to solve the above problems.

YOLO network version has been updated from 1.0 to the current 10.0; considering its model size, detection accuracy, and detection speed, YOLOv5 still has strong application value. Compared to other versions of YOLO, YOLOv5 weighs the detection performance and model size and still contains three main parts: backbone network, feature fusion network, and detection head [15], which is more adapted to small target detection, with greater speed and detection accuracy [25]. According to the number of model channels and convolution times, the YOLOv5 family can be divided into YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x, which are similar in structure but different in size. YOLOv5s, with the smallest number of model channels and convolution times, belong to the lightest class, which has a small amount of computation and thus is faster and more adaptable to mobile terminal devices [2]. Firstly, YOLOv5s performs Mosaic data enhancement and size scaling and other operations on the input images, i.e., four input photos are used to be spliced by random scaling, random cropping, and random arranging, which not only enriches the dataset to reduce the memory needs of the device, but also accelerates the training speed; secondly, preprocessing the data by using Focus to reduce the amount of computation and increase the speed, alternately using three Convolutional Block Layers (CBL) and Cross Stage Partial networks (CSP) for feature extraction, deepening the network and preventing model overfitting, this part is the core of the backbone network, fusing local and global features by Spatial Pyramid Pooling (SPP) and inputting the extracted features in different stages into the neck network; then, in the neck network, the Feature Pyramid Network(FPN) and the Path Aggregation Network (PAN) are utilized for enhanced feature fusion and extraction, fusing the features extracted at different stages of the backbone and those obtained from the detection network, to improve the robustness and generalization of the detection; finally, the feature *mAPs* outputted from the Neck network are convolved and predicted to derive the bounding boxes, categories and confidence levels of the detection targets at different scales [26].

YOLOv5s algorithm weighs the detection accuracy and detection speed compared to other algorithms, but the model parameters are more and cannot be embedded in mobile device terminals. In order to make up for the shortcomings of the above research, we take the YOLOv5s algorithm for improvement.

2. Algorithms Overview

In order to better adapt the model to the field of aero-engine blade crack detection, this paper proposes a detection algorithm based on MobileNet3, YOLOv5, and GSConv (hereinafter referred to as MobGSim-YOLO algorithm), the structure of which is shown in Figure 2. The K-means++ algorithm is introduced instead of Kmeans, and the anchor frame size is re-selected to improve the detection accuracy; the backbone network is replaced with the lightweight module MobileNet3 to realize the initial extraction of the dataset features. The neck part utilizes the lightweight convolution of GSConv instead of the ordinary convolution, and the activation function is replaced with the Hard Sigmoid to improve the multi-scale fusion and to enhance the small target feature extraction ability in the engine interior dark light, complex structure, noise, and the blade cracks are usually small, deep feature fusion is often not conducive to the extraction of small targets, so the SimAM attention mechanism is introduced in the part of the feature fusion to make the model pay more attention to small target features. Compared with the YOLOv5s algorithm, the model proposed in this thesis is more innovative, almost updating the whole framework and retaining only a few original modules, which substantially improves the detection speed and compresses the model size under the premise of guaranteeing the detection accuracy. Finally, the model MobGSim-YOLO used in this paper is proposed.

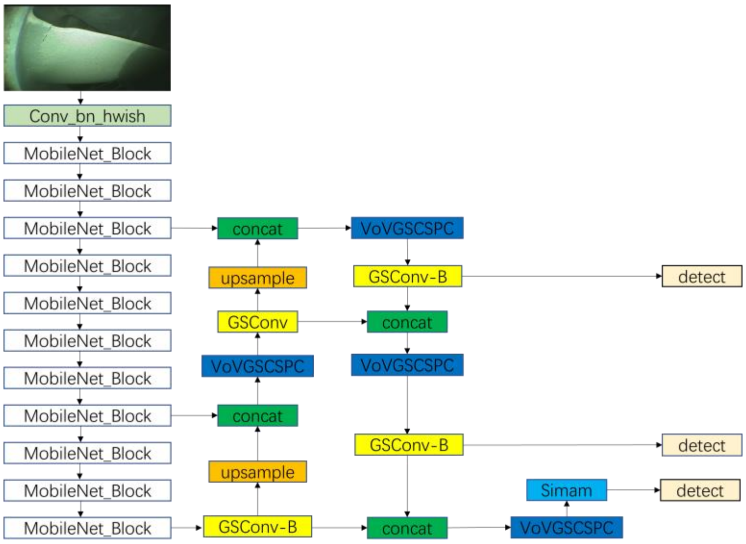


Figure 2. MobGSim-YOLO Algorithm Structure Diagram.

Target detection can be divided into two stages: one is to determine the target location, and the other is to identify the target category. For YOLOv5s, generating rectangular candidate frames or Kmeans clustering, mainly randomly selected *K* samples as the initial clustering center, cannot avoid the similarity problem of clustering, aiming frames to form the positional error affects the model accuracy, and the *K* value is difficult to determine.

Therefore, in this paper, the Kmeans++ algorithm is used to optimize the clustering center selection so as to get the most suitable aiming frame. The steps are as follows:

(1) Randomly select a sample from the data distribution set as the initial clustering center, as close to the edge as possible;
(2) Calculate the distance between each sample and the current clustering center and select the minimum distance of $D(x)$;
(3) Calculate the probability that the sample in (2) will be the next cluster center according to the probability distribution formula P. The formula is given as follows:

$$P = \frac{D(x)^2}{\sum\limits_{x \in X} D(x)^2} \tag{1}$$

Repeat the above (2) and (3) until K clustering centers are selected and iterate using the Kmeans++ algorithm to get the final result. As Table 1 shows the results of anchor frame generation, the comparison shows that the anchor frames obtained by K-means clustering have a deviation in the center and small size, while the center deviation obtained by Kmeans++ clustering is small and the size is appropriate, and most of the cracks can be included. w denotes the length of the anchor frame, and h denotes the width of the anchor frame. Therefore, the K-means clustering algorithm is more reliable in target detection.

Table 1. Anchor Box Generation Results.

Algorithm	Feel the Wild Size	Anchor Frame Size (w, h)
K-means	Large	(140, 98) (139, 161) (118, 205)
	Medium	(75, 36) (68, 71) (65, 96)
	Small	(9, 4) (12, 16) (11, 20)
Kmeans++	Large	(166, 129) (157, 196) (144, 255)
	Medium	(102, 69) (87, 91) (80, 123)
	Small	(16, 11) (22, 34) (20, 52)

The backbone network is an important part of the deep learning model, which mainly realizes the function of extracting features and learning them. The backbone network of YOLOv5s utilizes the complex C3 network structure, which is not only slow to detect but also has many parameters and large computations, which is demanding on the equipment, and it cannot be used in the borehole probing equipment. Intelligent borehole probing devices need to embed deep learning models into mobile device terminals with limited device memory and performance, which undoubtedly puts high demands on the computation and memory of the models. Several factors need to be considered to choose a suitable backbone network: image dataset characteristics and size, task-specific requirements, etc. With comprehensive consideration, we choose the more advanced MobileNetv3 model as the backbone network and improve it.

MobileNet network architecture is a lightweight neural network model proposed by Google, which has now been updated to the fourth generation. It relies on lightweight model design and is widely used in projects with limited hardware resources and arithmetic power. Compared to other versions, MobileNetv3 is applied to this dataset with better results, utilizing the inverted residual structure, depth separable convolution, SE attention mechanism, and activation function h-swish [1].

The principle of traditional multi-channel convolution is shown in Figure 3, the size of the input feature map is: $C_i \times H_i \times W_i$, the size of the convolution kernel is $C_i \times D_k \times D_k$, and the number is C_o. The original feature map is computed with each original convolution kernel for channel convolution respectively, and all the results can be summed up to get the output feature map, and the number of parameters is: $S = D_K \times D_K \times C_0 \times C_i$.

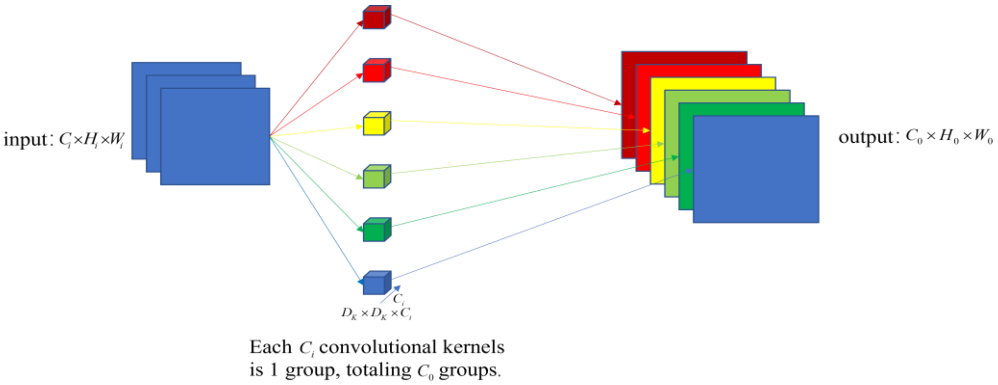

input: $C_i \times H_i \times W_i$
output: $C_0 \times H_0 \times W_0$
C_i
$D_K \times D_K \times C_i$
Each C_i convolutional kernels
is 1 group, totaling C_0 groups.

Figure 3. Traditional Convolution Schematic.

Depth separable convolution can be understood as a tandem operation of channel-by-channel convolution and point-by-point convolution and the working principle is shown in Figure 4. The size of the input feature map is $C_i \times H_i \times W_i$, and each channel is convolved channel by channel with a convolution to obtain the intermediate feature map. The intermediate feature map is then convolved point by point with a convolution kernel, and the result is summed up as the output feature 2, and the number of convolution kernels point by point is equal to the number of channels of the output feature map. The number of parameters is: $P = D_K \times D_K \times C_i + 1 \times 1 \times C_i \times C_0$.

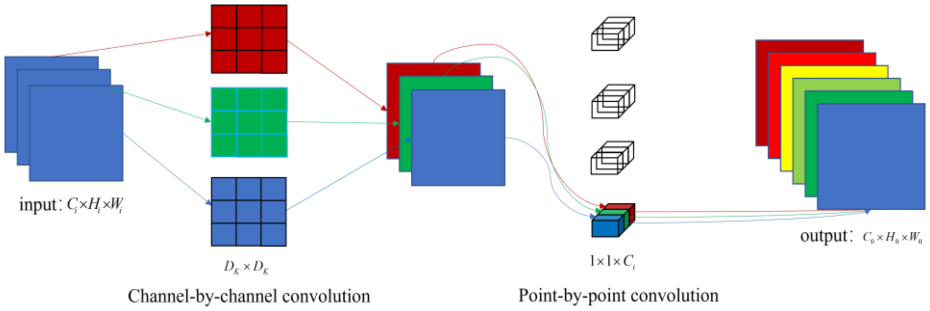

input: $C_i \times H_i \times W_i$
$D_K \times D_K$
$1 \times 1 \times C_i$
output: $C_0 \times H_0 \times W_0$
Channel-by-channel convolution
Point-by-point convolution

Figure 4. Depth separable Convolution Schematic.

The ratio of the number of parameters for conventional multichannel convolution and depth separable convolution is:

$$\frac{P}{S} = \frac{D_K \times D_K + C_0}{D_K \times D_K \times C_0} = \frac{1}{C_0} + \frac{1}{D_K \times D_K} \tag{2}$$

It can be seen that the depth separable convolution can make the model simpler and less parameterized, i.e., the depth separable convolution extracts deeper features with the same parameters. Specific references can be found in the literature [27].

The depth separable convolutional parameters are less, which will cause the model to lose a lot of key features; in order to ensure the accuracy of the model, MobileNet3 also introduces the SE attention mechanism. The SE attention mechanism mainly starts from the perspective of the channel to get the weights of different channels at different positions of the feature map and to learn the channel features more accurately. First of all, the feature map is globally average pooled and compressed into vectors, i.e., each channel can be represented by a single number; then the weights of each channel are generated by two fully connected layers; finally, the generated channel weights are used to assign values to the original feature map to get the final required feature map, and the details can be referred to the literature [28], and the schematic diagram is shown in Figure 5.

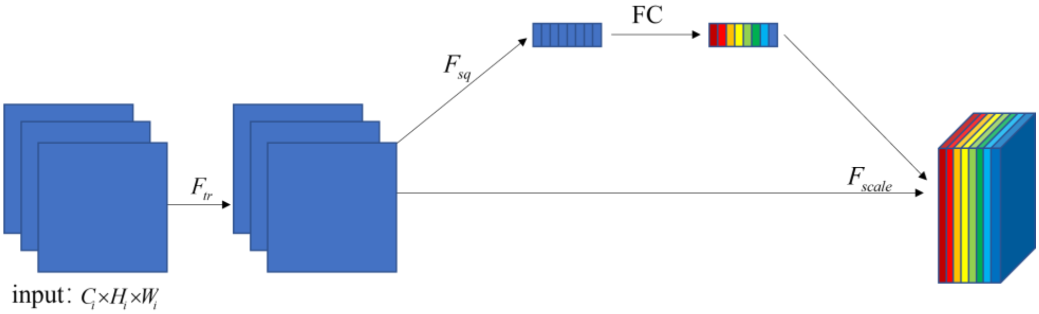


Figure 5. SE Attention Mechanism Schematic.

A reverse residual network replaces the traditional ResNet [13] residual structure. The reversed residual network first uses dot convolution to boost the number of channels of the feature map, uses depth-separable convolution to extract features in high-dimensional space, which reduces the parameters, and then uses dot convolution to reduce the number of channels, and introduces a new activation function, ReLU6, and the formula can be expressed as:

$$\mathrm{Re}LU6(x) = \min(\max(0, x), 6) \tag{3}$$

In addition to this, in order to ensure the nonlinearity, lightness, and accuracy of the activation function, MobileNet3 also introduces the h-swish function, which can be expressed as an equation:

$$h - swish(x) = x\frac{\mathrm{Re}LU6(x+3)}{6} \tag{4}$$

Traditional activation function has a strong exponential nature, and the derivation is complex and computationally intensive in the gradient calculation; whereas the *h-swish* function is a combined form of linear *ReLU* activation functions with at most quadratic terms, which is simple and smoother to compute, and reduces the computational cost significantly.

The smart hole detector device should be integrated into the mobile device terminal; the algorithm model size has high requirements, the ordinary convolution to extract features is limited, and for more parameters, we use the GSConv lightweight convolution module instead of the ordinary convolution and use the VOVGSCSPC module [29], i.e., slim-neck. The slim-neck module was first proposed to be applied in the automatic driving system, which not only requires high detection accuracy but also needs good real-time performance, which is basically the same as the requirements of this paper's smart hole detector. GSConv is more effective in lightweight networks than other convolutional modules, which combine the ideas of GhostNet and ShuffleNetv2 and apply deep separable convolution more skillfully. The principle of GhostConv is to use a small number of ordinary convolutions to compress the feature channels, which can reduce the amount of computation, and then carry out the constant to get the Intrinsic feature and, at the same time, extract the Ghost feature through the low-cost depth-separable convolution, and then finally splice the two types of feature to get the final feature. GSConv principle uses a smaller ordinary convolution to generate a series of basic features, and then a series of features are generated using depth separable convolution. Then, the two groups of features share weights and merge them, randomly mixing and washing the feature to enhance the generalization ability, to ensure that the information circulates between different groups, and then finally form the final convolutional results, it is obvious that GSConv has more outputs compared to the ordinary convolution, but the cost of computation is still kept at a low level, and its schematic diagram is shown in Figure 6.

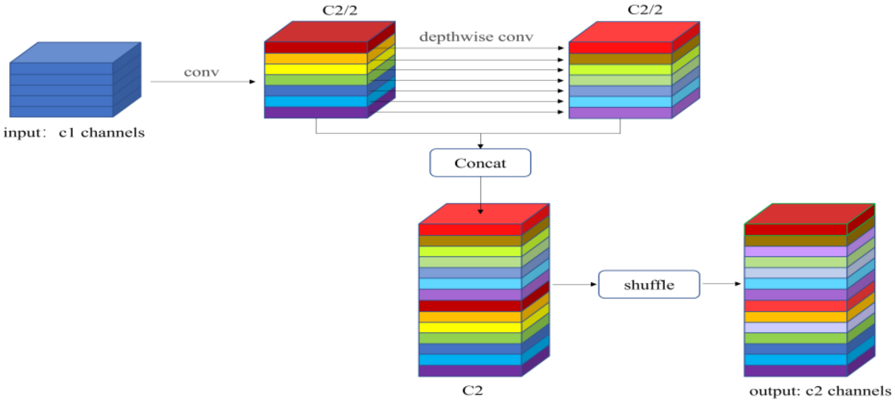


Figure 6. GSConv Schematic.

Two GSConv and one ordinary convolution are combined in parallel to generate the GS bottleneck, the structure of which is shown in Figure 7. This module is then combined with ordinary convolution to be the VOVGSCSPC module, which is obtained by ablation experiments and one-time aggregation methods, and its structure is shown in Figure 8. Research manuscripts reporting large datasets that are deposited in a publicly available database should specify where the data have been deposited and provide the relevant accession numbers. If the accession numbers have not yet been obtained at the time of submission, please state that they will be provided during review. They must be provided prior to publication. For details, please refer to the literature [29].

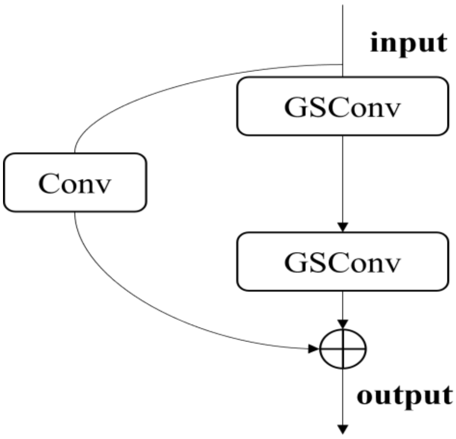


Figure 7. GS bottleneck.

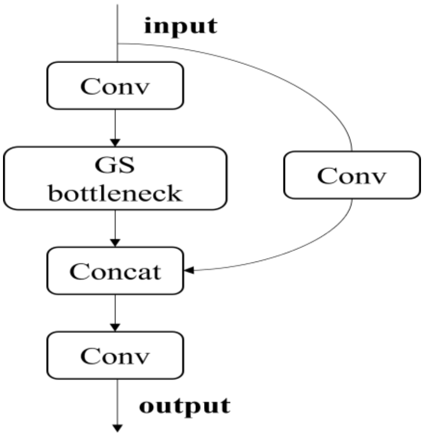


Figure 8. VOVGSCSPC.

The activation function for the shuffle blending operation in GSConv is a Sigmoid function. The Sigmoid function and derivatives are schematically shown in Figure 9. It can be seen that the two ends of the function are saturated regions, and the derivative of the saturated region is close to 0, which will bring serious gradient vanishing problems. When the network is deep enough, the gradient will gradually disappear, reducing the convergence speed of the network. In addition, the Sigmoid function is an exponential-type function, and the computational cost is too high. Although the slim-neck structure improves a lot in performance, the limited number of aero-engine blade datasets is prone to the risk of overfitting, for which this paper redesigns the slim-neck and optimizes the structure again.

In this paper, Hard Sigmoid function is used to replace the original activation function, and its core idea is to approximate the Sigmoid function by a segmented linear function with the expression:

$$f(x) = \min(\max(x+3),0)/6$$

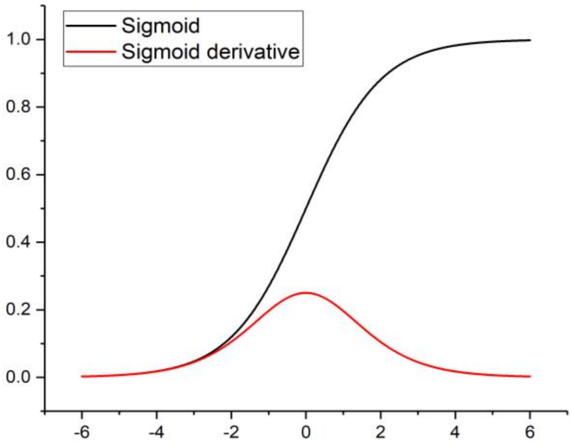


Figure 9. Sigmoid and derivative schematics.

A comparison of the Sigmoid function and Hard Sigmoid function is shown in Figure 10. Compared with the original activation function, the Hard Sigmoid function has a more stable gradient in the saturation region and is less prone to gradient vanishing. In order to verify whether replacing this activation function helps to improve the performance of the original model, ablation experiments are also done in this paper.

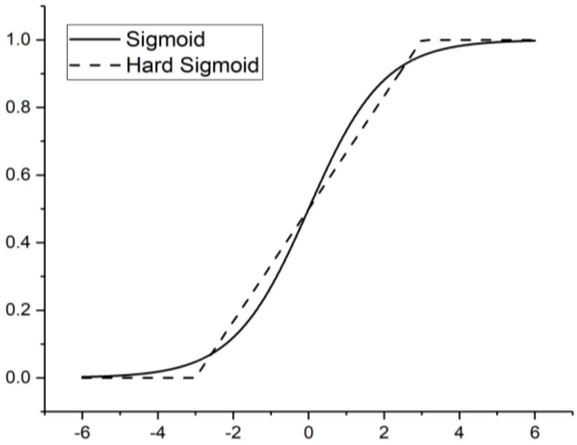


Figure 10. Sigmoid and Hard Sigmoid.

In the dark and noisy interior of an aero-engine, theoretically, a deeper model can learn a more complex feature representation, which is more conducive to distinguishing between noise and small targets, but at the same time, multiple down-sampling operations, with the increase of the network depth, lead to an increase in the receptive field, and the representation of the small targets on the feature will be better sparse or even lost. To overcome this difficulty, we introduce the attention mechanism to emphasize important regions. Learning the attention mechanism module is widely used in many computer vision projects because it allows the model to focus on more important information and ignore unimportant information. Aero-engine blade crack detection is different from traditional target detection in that blade cracks tend to be long, diversified in shape, and have inconspicuous features; secondly, the engine interior has low brightness, complex structure, less visible area, and more background noise. In order to overcome the influence of the above situations and considering the lightweight design, we adopt the SimAM attention module, as shown in Figure 11, which will reasonably assign 3D weights to the feature and increase the degree of target attention so that the model can compute the local similarity between the target region and the neighboring regions, capture the texture features of the image, and improve the recognition accuracy of the crack. Compared to the current hot attention mechanisms, including ECA, CBAM, SE, etc., the biggest advantage of SimAM is that it does not need to add parameters to the original network, and it is a lightweight module while still maintaining considerable accuracy [11]. This attention module mainly takes inspiration from human neurons and introduces an energy function to assign weights; the larger the energy difference between a neuron and its surrounding neurons, the more important and worthy of attention the neuron is. For details, please refer to the literature [11].

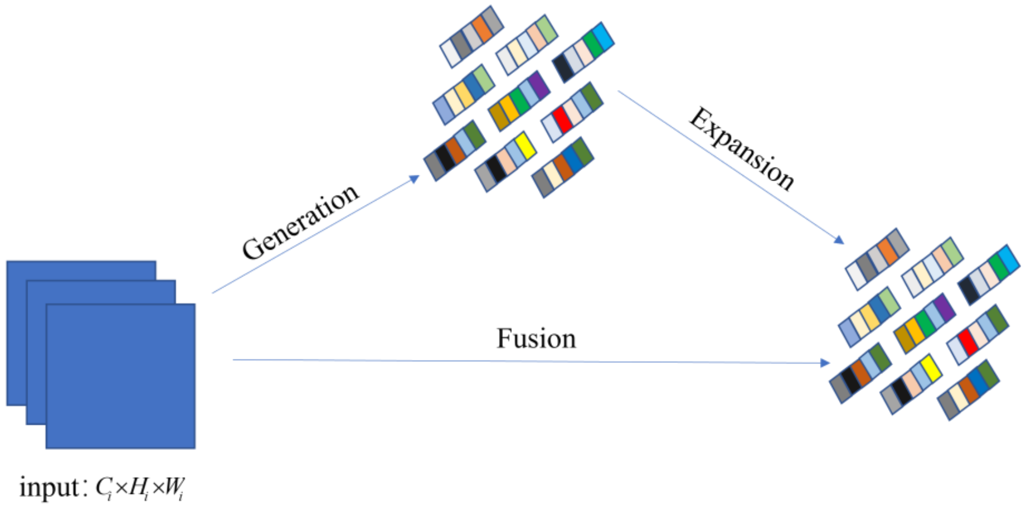


Figure 11. SimAM Schematic.

3. Experiments and Analysis of Results

The aero-engine blade crack intelligent detection process is shown in Figure 12. Firstly, the aero-engine blade with cracks is manually screened, and the dataset is expanded by cropping and rotating the images and divided into training, validation, and testing sets; secondly, all the image sizes are modified to be 640 × 640 × 3, and the usable aero-engine blade cracks dataset is obtained; the training parameters are set up, and the COCO dataset [30] is utilized to perform deep learning model preprocessing, and get the initial weights and biases; then adjust the training parameters, and use the obtained aero-engine blade crack dataset for freeze training, and the model weights and biases are updated; finally, validate and test the model.

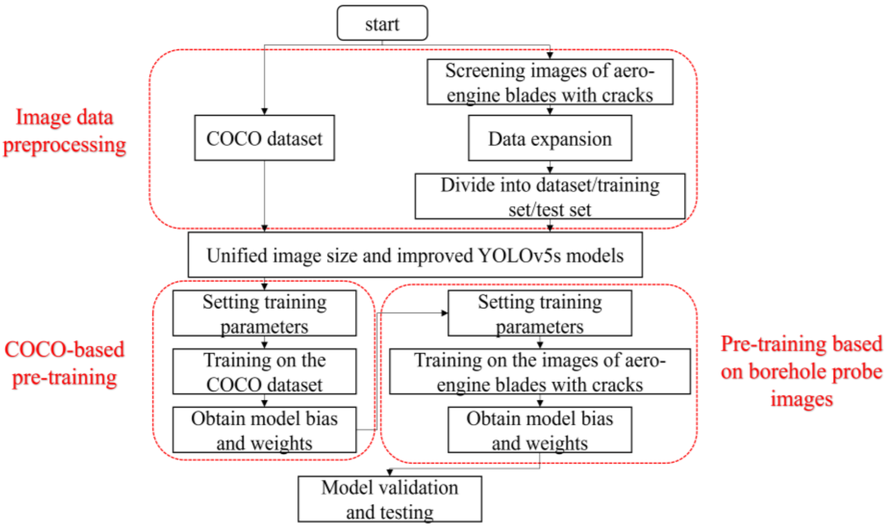

start
Image data preprocessing
COCO dataset
Screening images of aero-engine blades with cracks
Data expansion
Divide into dataset/training set/test set
Unified image size and improved YOLOv5s models
COCO-based pre-training
Setting training parameters
Training on the COCO dataset
Obtain model bias and weights
Setting training parameters
Training on the images of aero-engine blades with cracks
Obtain model bias and weights
Pre-training based on borehole probe images
Model validation and testing

Figure 12. The aero-engine blade cracks the intelligent detection process.

3.1. Data Set Production

The datasets explored in this study were derived exclusively from photographs captured by an aero-engine equipped with a borescope. Initially, aero-engine blade images exhibiting cracks were meticulously filtered out, totaling approximately 200. These images were subsequently labeled with the identifier "crack" utilizing specialist labeling software, ensuring accurate categorization, and the images are expanded, which are shown in Figure 13. and randomly divided into training set, validation set, and test set according to the ratio of 7:2:1, and the resolution of all the images is unified to be 640 × 640. A total of 3000 aero-engine blade images were obtained in the end. Crack images, the use of image annotation software, labeling sequentially on all cracks in accordance with the VOC data format, and finally, automatically recording the crack location and rectangular box size in a notepad file.

3.2. Model Evaluation Indicators

In order to accurately measure the performance of the model, this paper selects standard evaluation metrics for quantitative evaluation, such as Precision, Recall, mAP@0.5, mAP@0.95, parameter, FLOPs and Frames Per Second (FPS). The binary confusion matrix is widely used in evaluating target detection models, which simply means that the number of model predictions and the number of true labels for each category are integrated into a single matrix.

Precision is the ratio of the number of model-predicted positive samples that are also positive to the number of all model-predicted positive samples, and Recall is the ratio of the number of model-predicted positive samples that are also positive to the number of all actual positive samples. The formula is as follows:

$$P = \frac{TP}{TP + FP} \tag{5}$$

$$R = \frac{TP}{TP + FN} \tag{6}$$

where P denotes precision; R denotes Recall; TP denotes the number of samples predicted to be positive and actually also positive; FP denotes the number of samples predicted to be positive and actually negative; and FN denotes the number of samples predicted to be negative and actually positive.

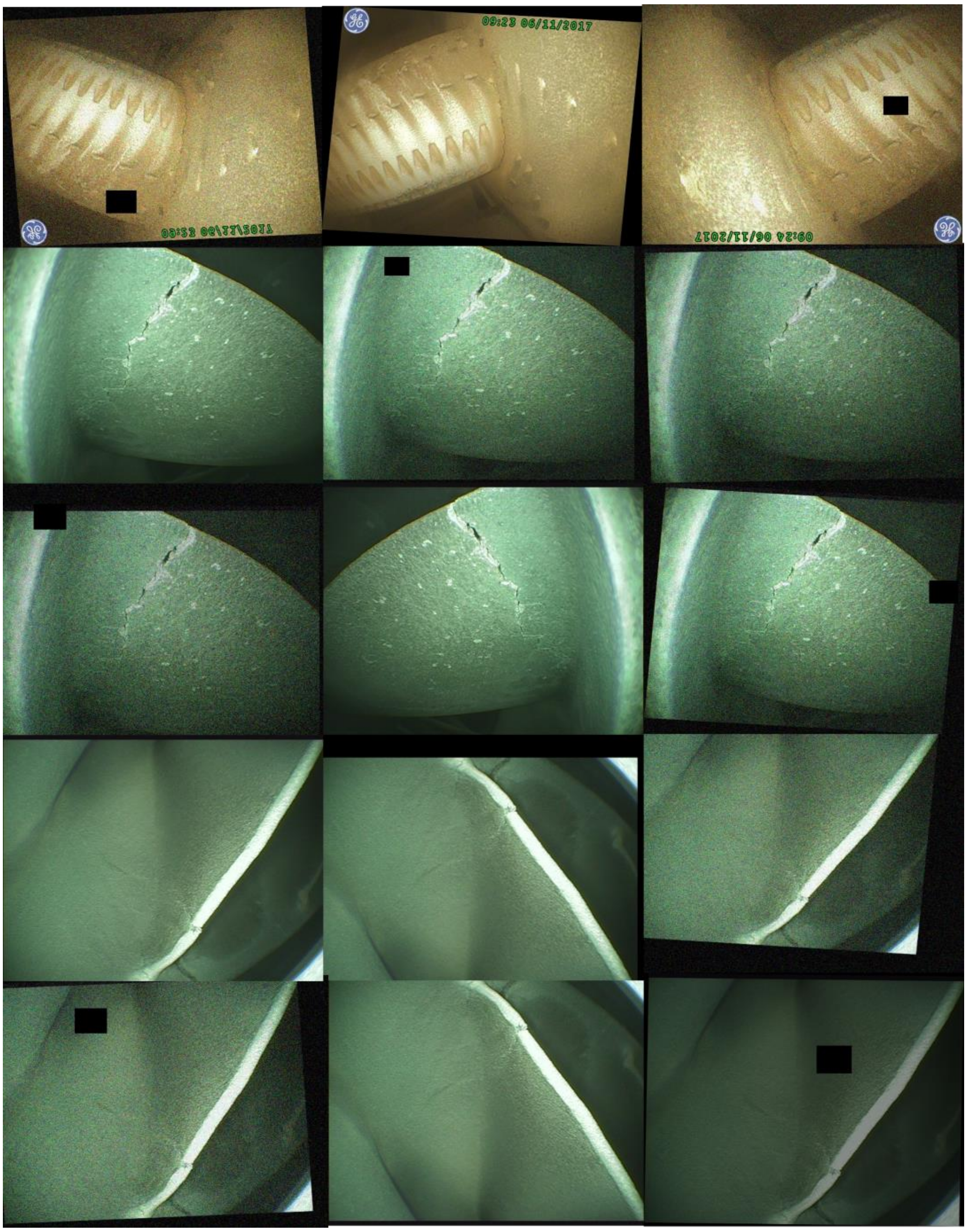


Figure 13. Image Expansion.

Recall for the horizontal axis, Precision for the vertical coordinates of the curve called *P-R* curve, the curve and the axis of the polygon around the city of the area known as the *AP*, the average of the *AP* can be obtained by averaging the mAP.

The formula for mAP is:

$$mAP = \frac{1}{K}\sum_{i=1}^{K} AP_i \tag{7}$$

K is the number of detection categories, which is 1 in this paper.

To better illustrate mAP@0.5 and mAP@0.95, the concept of *IOU* must be introduced here, which measures the degree of overlap between the predicted bounding box and the true bounding box, the intersection is the area where the predicted box and the true box overlap, the concatenation is the sum of the areas of all the regions of the predicted box and the true box, and the *IOU* is the ratio of the intersection and the concatenation. The schematic diagram is shown in Figure 14. S1 denotes the area of the region where the prediction box and the real box overlap; S2 denotes the sum of all areas of the prediction box and the real box. The formula for *IOU* is:

$$IOU = \frac{S_1}{S_2} \tag{8}$$

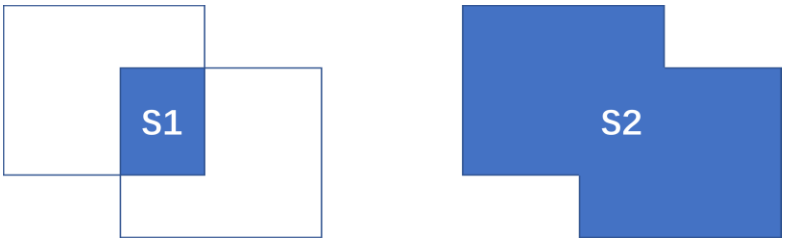


Figure 14. IOU Explanatory Chart.

mAP@0.5 then denotes the average precision mean *mAP* of $IOU \geq 0.5$. And mAP @0.95 is stricter, denoting the average precision mean of $IOU \geq 0.95$.

All of the above are measures of model detection accuracy; the smart hole detection device also requires the model to be small enough and the detection speed to be fast enough. Therefore, it is also necessary to introduce the model size (parameter), FLOPS, and FPS. Parameters include weights, neuron paranoia, convolution kernel, all-connected layer weights, anchor frame parameters, etc. FLOPS floating point operations per second characterize the computational power; FPS can reflect the speed of the model testing and inference, i.e., how many frames of images can be processed per second, which is extremely important for real-time monitoring of the borehole detection equipment. This is extremely important for real-time monitoring of equipment.

3.3. Ablation Experiment

In this paper, we use the Windows 10 operating system, the CPU is Intel(R)Core(TM)i5-10200H, and the GPU is NVIDIA GeForce GTX 1650 Ti. YOLOv5s model is built based on Pytorch deep learning framework using Python 3.6 programming language. The model training parameters: epochs are 300, batch size is 2, image size is 640, initial learning rate is 0.01, momentum parameter is 0.937, weight decay coefficient is 0.0005, and SGD optimizer is used to iterate the parameters.

3.4. Ablation Experiment

In order to prove the effectiveness of the proposed algorithm in this paper, we perform ablation experiments on the same engine blade crack dataset and ensure that the training strategy and hyperparameters are the same. The design of the ablation experiment is as follows:

- Module A1: Introducing the K-means++ clustering algorithm;
- Module A2: Replace the backbone network with MobileNet3 for a lightweight design for downsizing the model and reducing parameters;
- Module A3: Replace the neck part of the ordinary convolution as GSConv and add the VOVGSCSPC module after it to form a slim-neck module, which enhances the depth of the network and distinguishes between noise and valid features;
- Module A4: Slim-neck module after replacing activation function by hard Sigmoid;
- Module A5: Incorporating SimAM attention mechanisms to improve the learning of small features.

The experimental results are shown in Tables 2–4:

Table 2. YOLOv3, YOLOv4, and YOLOv7 test results comparison chart.

	Precision	Recall	mAP@0.5	mAP@0.95	Parameter (MB)	FLOPS	FPS
Yolov3	0.71	0.57	0.64	0.334	28.76	16.8	60.5
Yolov4	0.72	0.579	0.655	0.346	27.64	16.5	61.8
Yolov7	0.821	0.697	0.758	0.424	25.45	15.2	65.7

Table 3. Comparison of test accuracy for ablation experiment 1. (√ denotes the module in which the model exists, All of the following.)

Yolov5s	A1	A2	A3	A4	A5	Precision	Recall	mAP@0.5	mAP@0.95
√						0.72	0.61	0.67	0.352
√	√					0.793	0.565	0.668	0.356
√	√	√				0.811	0.708	0.769	0.414
√	√	√	√			0.87	0.778	0.832	0.507
√	√	√	√		√	0.892	0.808	0.848	0.518
√	√	√		√	√	0.931	0.792	0.849	0.528

Table 4. Comparison of model size and test speed for ablation experiment 1.

Yolov5s	A1	A2	A3	A4	A5	Parameter (MB)	FLOPS	FPS
√						26.75	15.8	72.4
√	√					26.75	15.8	72.3
√	√	√				13.5	6.2	99.1
√	√	√	√			12.65	4.2	99
√	√	√	√		√	12.65	4.2	100.3
√	√	√		√	√	12.58	4.2	95.8

By analyzing the results of the experiment, the following conclusions can be drawn:

1. The introduction of K-means++ clustering frames made the generated anchor frames more adapted to the present data, which significantly improved the test accuracy, and the results were in line with the expected assumptions;
2. Replacing the backbone network with the lightweight module MobileNet3, the model parameters are reduced by 49.54% compared to YOLOv5s, and the model accuracy is also slightly improved, mainly due to the small number of parameters in depth separable convolution compared to the ordinary convolution, the small number of parameters in h-swish activation function compared to the exponential activation function, the SE attention mechanism to improve the attention to the features, and the inverted residual structure enhances the model expression ability, and the results are in line with the expected assumptions;
3. Replacing the ordinary convolution in the neck with GSConv and adding the VOVGSC-SPC module, i.e., slim-neck module, afterward, the results show that compared with YOLOv5s, the parameters are reduced by 52.71%, and the accuracy is improved by 20.83%, which is significant due to the fact that the model learns deeper features after the introduction of slim-neck, which enhances the feature fusion and improves the target learning ability, which is in line with the expected conception;
4. Adding the SimAM attention mechanism to the head’s front-end, which does not introduce any parametric quantities but improves the accuracy, can improve the problem of deeper learning depth and ignoring small target features brought about by GSConv, which is in line with the expected conception.
5. After replacing the activation function with GSConv, the results compared with YOLOv5s, the parameters were reduced by 52.96%, and the accuracy was improved by 29.31%, which was obvious and in line with the expected assumptions.
6. After the lightweight design, the FPS of all the models is greater than 95, which meets the detection requirements of the hole detector.

There are many popular lightweight backbone networks, such as PP-LCNet and MobileNet series. In order to enhance the persuasiveness and visualize the results, this paper also conducts a series of ablation experiments for backbone networks. The second set of ablation experiments is designed as follows:

- Module A6: Replace the backbone network with PP-LC [31];
- Module A7: Replace the backbone network with MobileNetv4 [32].

The experimental results are shown in Tables 5 and 6:

Table 5. Comparison of test accuracy for ablation experiment 2.

Yolov5s	A1	A2	A3	A4	A5	A6	A7	Precision	Recall	mAP@0.5	mAP@0.95
√	√	√		√	√			0.931	0.792	0.849	0.528
√	√			√	√	√		0.813	0.705	0.775	0.425
√	√			√	√		√	0.78	0.68	0.755	0.388

Table 6. Comparison of test accuracy for ablation experiment 2.

Yolov5s	A1	A2	A3	A4	A5	A6	A7	Parameter (MB)	FLOPS	FPS
√	√	√		√	√			12.65	4.2	100.3
√	√			√	√	√		12.39	4.3	99.5
√	√			√	√		√	12.51	4.4	145.4

By analyzing the results of the second group of ablation experiments, the model that replaces the backbone network for PP-LC and MobileNet4, although excellent in model size and detection speed, the decline in test accuracy is more obvious and can't meet the accuracy requirements, so by analyzing the experimental data, we can get the best performance of the backbone network for MobileNet3, which is in line with the application scenario of this paper.

Through the above ablation experiments, it can be concluded that the model, after improving on the YOLOv5s model by adding the K-means++ algorithm, replacing the backbone network with the Mobilenetv3 lightweight module, replacing the neck portion with the slim-neck structure, and incorporating the SimAM attention mechanism performs the best when weighed against the test accuracy, speed, and model size. The training results are shown in Figure 15. Analyzing the loss function curve shows that the total loss value of the model decreases with the increase of training rounds, indicating that the model is gradually learning and reducing the prediction error, which is in line with the expected results. The model only detects one type of crack, so the classification damage is 0. Analyzing the precision, recall, and map curves, it can be seen that with the increase of training rounds, the performance index rises significantly and gradually reaches convergence after 300 rounds of training, and the training results are real and credible.

The test set is detected using the model MobGSim-YOLO algorithm proposed in this paper, and some of the plots of the results are shown in Figure 16. The YOLOv5s algorithm detection results are shown in Figure 17. The test selected in this paper is composed of photos of compressor blades and turbine blades, totaling 300, and only 7 turbine blade cracks and 5 compressor crack detection results are screened and displayed for the convenience of comparison. It can be observed by comparison that the confidence level of the MobGSim-YOLO model test results is significantly higher than that of the YOLOv5s model, visually observing that the bounding box formed by the former is more accurate, while the latter omits the boundary cracks with darker brightness and specific angles and misjudges the boundaries of the occluded and darker parts as cracks. It is clear that the MobGSim-YOLO model is superior in detection.

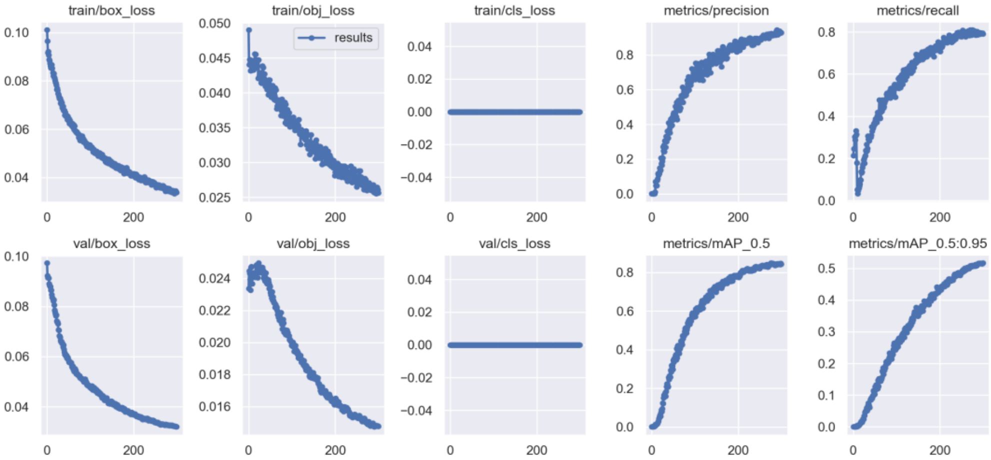


Figure 15. MobGSim-YOLO algorithm training results.

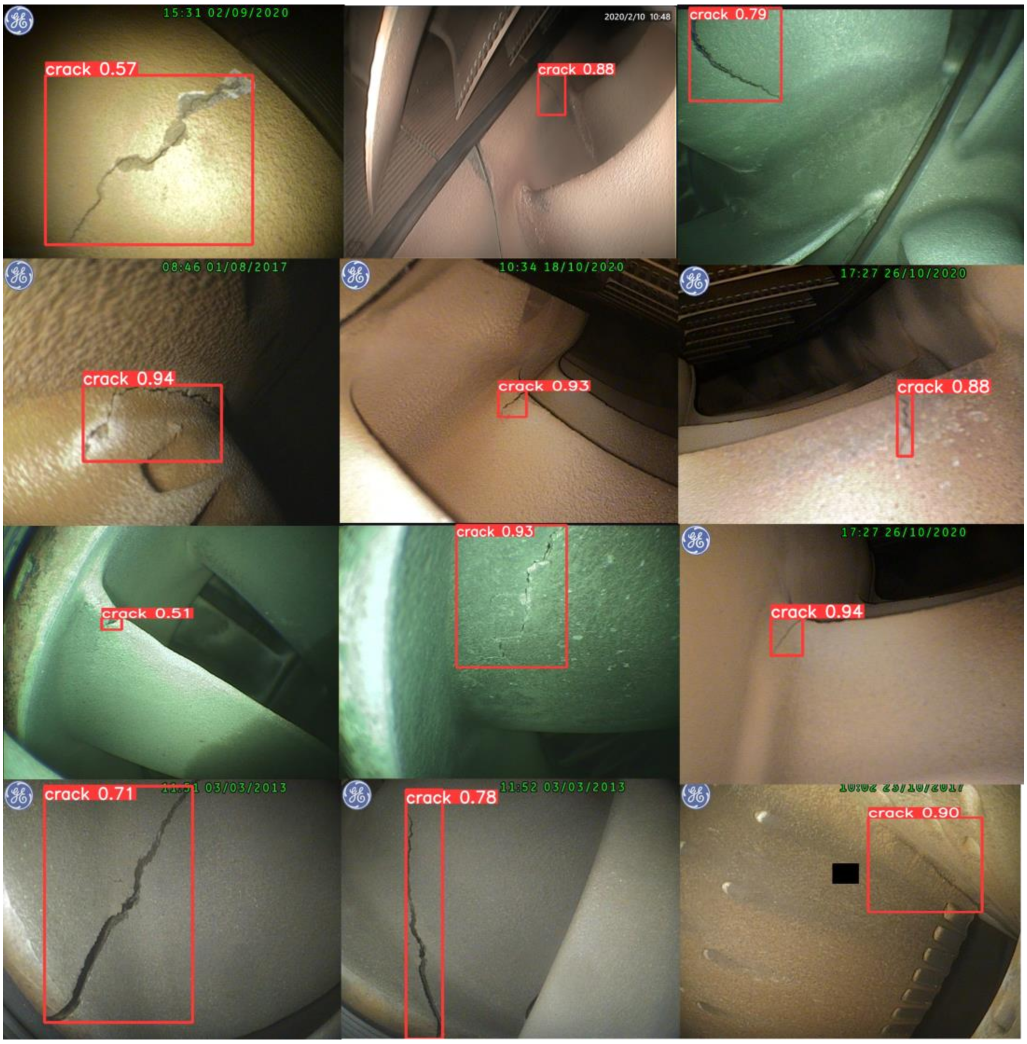


Figure 16. Test results for the MobGSim-YOLO model.

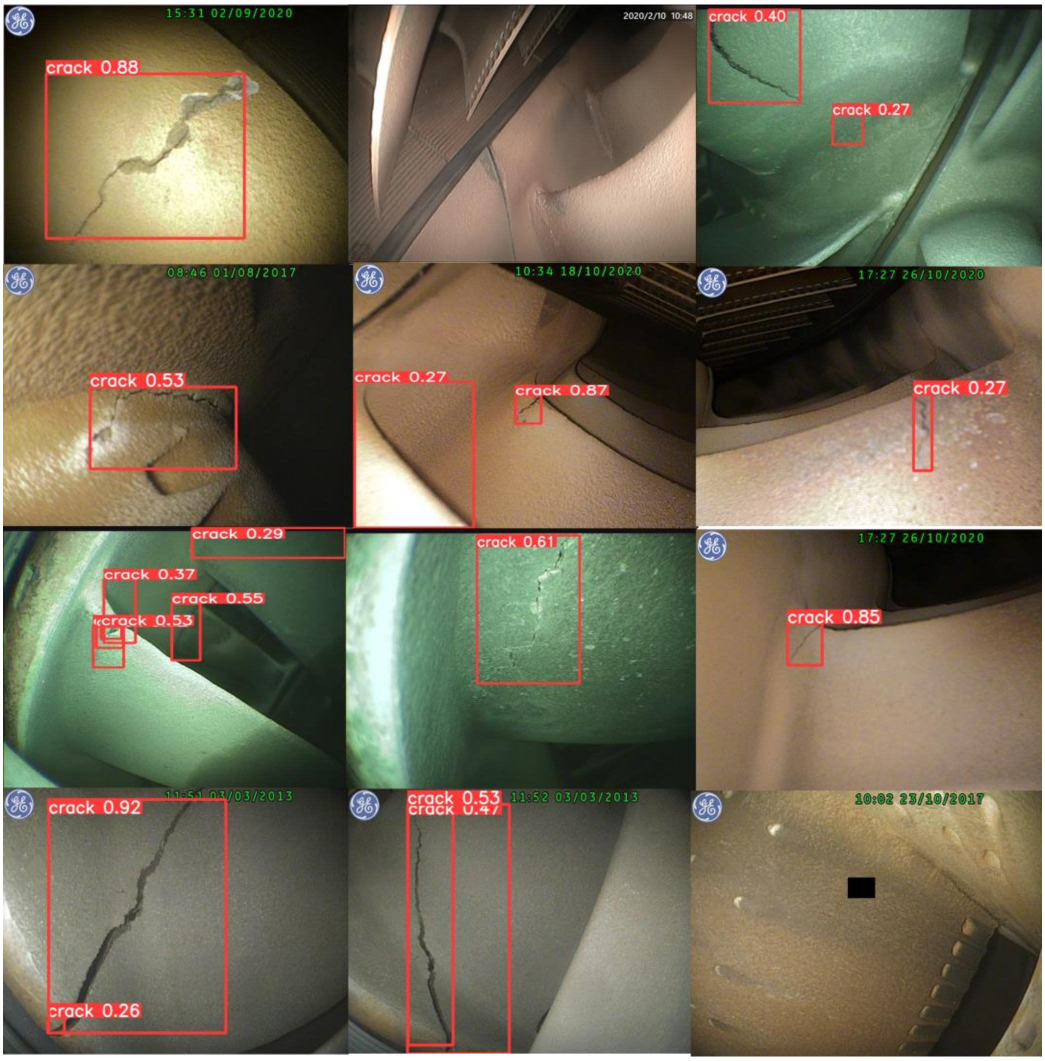


Figure 17. Test results for the YOLOv5s model.

4. Conclusions

In this paper, the MobGSim-YOLO model for hole detection is proposed for aero-engine blade crack detection. In this paper, we first summarize the problems of the current aero-engine intelligent hole detection, especially the lower detection accuracy, slower detection speed, and the larger model that cannot be embedded into the mobile device terminal. In order to improve the accuracy of YOLOv5s model, we utilize kmeans++ to reasonably design the clustering frame to improve the accuracy of the detection location, introduce and redesign the slim-neck module to improve the model depth and accuracy while ensuring that the size of the model is not increased, which is effective; embed the SimAM attention mechanism in the front-end of the head, to overcome the problem of small target loss due to the excessive depth of the model. We also embed the SimAM attention mechanism in the front of the head to overcome the problem of small target loss due to the deep model. We also replaced the backbone network with a Mobilenetv3 module for lightweight design. In this paper, a large number of ablation experiments are carried out on the real aero-engine blade crack dataset to verify the feasibility and accuracy of the method, and the result proves that the model can weigh the test accuracy and model size and fully comply with the requirements of the intelligent borehole detection equipment, which is an attempt to intelligent borehole detection of the aero-engine, and we hope that the model proposed in this paper can bring certain thinking.

Author Contributions: Conceptualization, X.H. and H.Z.; methodology, X.H.; software, W.W.; validation, L.J., J.P. and W.W.; formal analysis, L.J.; investigation, J.P.; resources, H.Z.; data curation, H.Z.; writing—original draft preparation, X.H.; writing—review and editing, X.H.; visualization, L.J.; supervision, H.Z.; project administration, J.P.; funding acquisition, H.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Zhao, L.; Wang, L. A new lightweight network based on MobileNetV3. *KSII Trans. Internet Inf. Syst.* **2022**, *16*, 1–15.
2. Yang, J.; Liu, S.; Li, Z.; Li, X.; Sun, J. Real-time object detection for streaming perception. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 5385–5395.
3. Shang, H.; Sun, C.; Liu, J.; Chen, X.; Yan, R. Deep learning-based borescope image processing for aero-engine blade in-situ damage detection. *Aerosp. Sci. Technol.* **2022**, *123*, 107473. [CrossRef]
4. Fahr, A. *Aeronautical Applications of Non-Destructive Testing*; DEStech Publications, Inc: Lancaster, PA, USA, 2013.
5. Wang, P.; Wang, W.; Zheng, S.; Chen, B.; Gao, Z. Fatigue Damage Evaluation of Compressor Blade Based on Nonlinear Ultrasonic Nondestructive Testing. *J. Mar. Sci. Eng.* **2021**, *9*, 1358. [CrossRef]
6. Xie, R.; Chen, D.; Pan, M.; Tian, W.; Wu, X.; Zhou, W.; Tang, Y. Fatigue crack length sizing using a novel flexible eddy current sensor array. *Sensors* **2015**, *15*, 32138–32151. [CrossRef]
7. Yang, G.; Wang, C.; Zhang, Y.; Song, K. Test Study of Automatic Eddy Current Detection of Aero-engine Turbine Blade. *Fail. Anal. Prev.* **2022**, *17*, 334–339.
8. Liu, L.; Yu, H.; Zheng, C.; Ye, D.; He, W.; Wang, S.; Li, J.; Wu, L.; Zhang, Y.; Xie, J. Nondestructive thickness measurement of thermal barrier coatings for turbine blades by terahertz time domain spectroscopy. *Photonics* **2023**, *10*, 105. [CrossRef]
9. Kutman, M.K.; Muftuler, F.B.; Harmansah, C.; Guldu, O.K. Use of bacteria as fluorescent penetrant for penetrant testing (PT). *J. Nondestruct. Eval.* **2020**, *39*, 15. [CrossRef]
10. Yu, H. Borescope and its application in aero engine maintenance. *Aeronaut. Manuf. Technol.* **2005**, *99*, 94–96.
11. Yang, L.; Zhang, R.-Y.; Li, L.; Xie, X. Simam: A simple, parameter-free attention module for convolutional neural networks. In Proceedings of the International Conference on Machine Learning, Virtual, 18–24 July 2021; pp. 11863–11874.
12. Wang, M.; Li, K.; Zhu, X.; Zhao, Y. Detection of surface defects on railway tracks based on deep learning. *IEEE Access* **2022**, *10*, 126451–126465. [CrossRef]
13. Wightman, R.; Touvron, H.; Jégou, H. Resnet strikes back: An improved training procedure in timm. *arXiv* **2021**, arXiv:2110.00476.
14. Bharati, P.; Pramanik, A. Deep learning techniques—R-CNN to mask R-CNN: A survey. In *Computational Intelligence in Pattern Recognition: Proceedings of CIPR*; Springer: Singapore, 2020; pp. 657–668.
15. Jiang, P.; Ergu, D.; Liu, F.; Cai, Y.; Ma, B. A Review of Yolo algorithm developments. *Procedia Comput. Sci.* **2022**, *199*, 1066–1073. [CrossRef]
16. Dai, J.; Li, Y.; He, K.; Sun, J. R-fcn: Object detection via region-based fully convolutional networks. In *Advances in Neural Information Processing Systems*; Tsinghua University: Beijing, China, 2016; Volume 29.
17. Wang, S.; Sun, G.; Zheng, B.; Du, Y. A crop image segmentation and extraction algorithm based on mask RCNN. *Entropy* **2021**, *23*, 1160. [CrossRef]
18. Upadhyay, A.; Li, J.; King, S.; Addepalli, S. A Deep-Learning-Based Approach for Aircraft Engine Defect Detection. *Machines* **2023**, *11*, 192. [CrossRef]
19. Li, D.; Li, Y.; Xie, Q.; Wu, Y.; Yu, Z.; Wang, J. Tiny defect detection in high-resolution aero-engine blade images via a coarse-to-fine framework. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 1–12. [CrossRef]
20. Zhang, D.; Zeng, N.; Lin, L. Detection of blades damages in aero engine. In Proceedings of the 2020 Chinese Automation Congress (CAC), Shanghai, China, 29 January 2021; pp. 6129–6134.
21. Li, B.; Wang, C.; Ding, X.; Ju, H.; Guo, Z.; Li, Z. Surface defect detection algorithm based on improved YOLOv4. *J. Beijing Univ. Aeronaut. Astronaut.* **2023**, *49*, 710–717.
22. Li, X.; Wang, W.; Sun, L.; Hu, B.; Zhu, L.; Zhang, J. Deep learning-based defects detection of certain aero-engine blades and vanes with DDSC-YOLOv5s. *Sci. Rep.* **2022**, *12*, 13067. [CrossRef] [PubMed]
23. Cai, S.; He, C. A damage detection method for aero-engine based on FDG-YOLO lightweight. *Model J. Beijing Univ. Aeronaut. Astronaut.* **2024**, 1–11. [CrossRef]
24. Li, S.; Yu, J.; Wang, H. Damages detection of aeroengine blades via deep learning algorithms. *IEEE Trans. Instrum. Meas.* **2023**, *72*, 5009111. [CrossRef]
25. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 779–788.

26. Jocher, G.; Chaurasia, A.; Stoken, A.; Borovec, J.; Kwon, Y.; Michael, K.; Fang, J.; Wong, C.; Yifu, Z.; Montes, D.J.Z. ultralytics/yolov5: v6. 2-yolov5 classification models, apple m1, reproducibility, clearml and deci. ai integrations. *Zenodo* **2022**. Available online: https://ui.adsabs.harvard.edu/abs/2022zndo...7002879J/abstract (accessed on 3 July 2024).
27. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv* **2017**, arXiv:1704.04861.
28. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141.
29. Li, H.; Li, J.; Wei, H.; Liu, Z.; Zhan, Z.; Ren, Q. Slim-neck by GSConv: A better design paradigm of detector architectures for autonomous vehicles. *arXiv* **2022**, arXiv:2206.02424.
30. Lin, T.-Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft coco: Common objects in context. In Proceedings of the Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, 6–12 September 2014; pp. 740–755.
31. Cui, C.; Gao, T.; Wei, S.; Du, Y.; Guo, R.; Dong, S.; Lu, B.; Zhou, Y.; Lv, X.; Liu, Q. PP-LCNet: A lightweight CPU convolutional neural network. *arXiv* **2021**, arXiv:2109.15099.
32. Qin, D.; Leichner, C.; Delakis, M.; Fornoni, M.; Luo, S.; Yang, F.; Wang, W.; Banbury, C.; Ye, C.; Akin, B. MobileNetV4-Universal Models for the Mobile Ecosystem. *arXiv* **2024**, arXiv:2404.10518.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.