

Article

SAB-YOLOv5: An Improved YOLOv5 Model for Permanent Magnetic Ferrite Magnet Rotor Detection

Bo Yu ^{1,2}, Qi Li ³, Wenhua Jiao ^{1,2,*}, Shiyang Zhang ^{1,2} and Yongjun Zhu ^{1,2}

¹ Artificial Intelligence Research Institute, China University of Mining and Technology, Xuzhou 221116, China; lb22060011@cumt.edu.cn (B.Y.); yj.zhu@cumt.edu.cn (Y.Z.)

² College of Information and Control Engineering, China University of Mining and Technology, Xuzhou 221116, China

³ School of Computer and Communication Engineering, Nanjing Tech University Pujiang Institute, Nanjing 211112, China

* Correspondence: wjiao@cumt.edu.cn

Abstract: Surface defects on the permanent magnetic ferrite magnet rotor are the primary cause for the decline in performance and safety hazards in permanent magnet motors. Machine-vision methods offer the possibility to identify defects automatically. In response to the challenges in the permanent magnetic ferrite magnet rotor, this study proposes an improved You Only Look Once (YOLO) algorithm named SAB-YOLOv5. Utilizing a line-scan camera, images capturing the complete surface of a general object are obtained, and a dataset containing surface defects is constructed. Simultaneously, an improved YOLOv5-based surface defect algorithm is introduced. Firstly, the algorithm enhances the capability to extract features at different scales by incorporating the Atrous Spatial Pyramid Pooling (ASPP) structure. Then, the fusion of features is improved by combining the tensor concatenation operation of the feature-melting network with the Bidirectional Feature Pyramid Network (BiFPN) structure. Finally, the introduction of the spatial pyramid dilated (SPD) convolutional structure into the backbone network and output end enhances the detection performance for minute defects on the target surface. In the study, the SAB-YOLOv5 algorithm shows an obvious increase from 84.2% to 98.3% in the mean average precision (mAP) compared to that of the original YOLOv5 algorithm. The results demonstrate that the data acquisition method and detection algorithm designed in this paper effectively enhance the efficiency of defect detection permanent magnetic ferrite magnet rotors.

Keywords: permanent magnetic ferrite magnet rotors; SAB-YOLOv5; YOLOv5; defect detection; deep learning; small object detection; attention mechanism

MSC: 68T07



Citation: Yu, B.; Li, Q.; Jiao, W.; Zhang, S.; Zhu, Y. SAB-YOLOv5: An Improved YOLOv5 Model for Permanent Magnetic Ferrite Magnet Rotor Detection. *Mathematics* **2024**, *12*, 957. <https://doi.org/10.3390/math12070957>

Academic Editor: Jonathan Blackledge

Received: 29 February 2024

Revised: 10 March 2024

Accepted: 13 March 2024

Published: 23 March 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In the production process of permanent magnetic ferrite magnet rotors [1], various manufacturing processes such as extrusion, molding, and sintering, as well as machining operations including turning, milling, and grinding can significantly impact quality [2].

The quality, in turn, directly influences the performance and lifespan of the permanent magnetic ferrite magnet rotor. Therefore, enhancing the production quality is of paramount importance [3]. In the early stages, defect detection in permanent magnetic ferrite magnet rotors primarily relied on manual visual inspection, as shown in Figure 1, which exhibited notable drawbacks [4]. This method heavily depended on the experience of operators, resulting in low detection efficiency, substantial manpower requirements, high workload, and relatively high manual inspection costs. Defects on the surface of magnetic rotors vary in scale and have low resolution, falling within the category of small targets [5]. This paper

investigates defect detection in magnetic rotor products, highlighting the prevalent issue of numerous small target defects in the products.

Traditional methods no longer suffice to meet the growing demands for defect detection in permanent magnetic ferrite magnet rotors [6]. Conventional target detection methods face challenges such as window redundancy, high time complexity, and a lack of purposeful region selection, which, in turn, limit detection accuracy. With the rapid development of convolutional neural networks, defect detection solutions based on deep learning have gradually found application in surface defect detection [7]. Researchers have discovered issues with traditional Gabor wavelet filters [8], namely a high parameter count and a slow detection speed of the algorithm. To address these issues, they proposed a composite differential evolution optimization method for Gabor filters. However, despite some progress in defect detection, this method still faces limitations in real-time detection. A method was proposed that determines thresholds and further segments defects through histogram analysis of defect and background regions [9]. Nevertheless, this method requires significantly distinguishable features in defect and background regions.

Over time, as convolutional neural networks have evolved, defect detection methods based on machine vision have emerged. An adaptive sorting fused attention network capable of selectively extracting visual features to enhance small target defect detection has been proposed [10]. However, this method exhibits suboptimal performance in the detection of a single type of defect. An attention cascade network defect detection algorithm was proposed, gradually increasing the detection heads by cascading two Intersection over Union (IoU) thresholds to improve defect detection performance [11]. However, this algorithm has limitations in modular network deployment. The You Only Look Once (YOLO) algorithm has been introduced, which simultaneously predicts bounding boxes and categories using a single convolutional neural network, enabling real-time processing of entire images [12]. Although the initial version of YOLO had shortcomings in accuracy and small target detection capabilities [13], it has been widely applied in the industrial sector with continuous improvements [14]. YOLOv1 addressed candidate region issues by utilizing global image information [15]. Moreover, YOLOv1 employed a single network to simultaneously learn detection and classification tasks in end-to-end training. Subsequently, YOLOv2 introduced batch normalization to reduce the covariate shift and used anchor boxes to handle targets of different sizes [16]. Additionally, YOLOv2 utilized high-resolution feature maps for detection to minimize object detection localization errors [17]. YOLOv3 introduced the Feature Pyramid Network (FPN) for multi-scale feature fusion and employed three different sizes of fully convolutional networks to predict targets of varying sizes [18]. Furthermore, YOLOv3 enhanced training data with operations such as random rotation and cropping for better adaptation to different scenarios. YOLOv4 employed techniques such as Spatial Pyramid Pooling [19] and a Path Aggregation Network for multi-scale feature fusion [20]. Additionally, YOLOv4 introduced the Mish activation function for better robustness and stability [21], along with improvements such as applying Mosaic data augmentation at the input end and utilizing DropBlock regularization [22]. YOLOv5 demonstrates a lower parameter count and lighter weight compared to YOLOX [23], while the performance of YOLOv8 [24] on specific tasks remains unverified. YOLOv5 has been verified to surpass other models in terms of generalization performance.



Figure 1. Manual visual inspection.

This study extracts, processes, and elaborates on information from target images with notable characteristics such as speed, high accuracy, and high repeatability. It not only enhances product quality, but also improves production efficiency and reduces the number of defective products, thus lowering the costs. With the ongoing advancement of networks, the accuracy of first-stage detection algorithms has steadily improved. From the aforementioned research, it can be inferred that the primary challenge for current defect detection methods for mechanical products lies in the trade-off between detection accuracy and speed. To achieve real-time detection while enhancing accuracy, especially for small targets, our focus was on the YOLOv5 object detection algorithm based on deep neural networks, combined with the improved methods such as mixed structures including ASPP, BiFPN, the SPD network in YOLOv5, and the proposed SAB-YOLOv5 method for permanent magnetic ferrite magnet rotors.

Based on the research findings in the domain of object detection algorithms, the main contributions and novel aspects of this research can be summarized as follows:

- A machine vision-based method for automatic identification of surface defects using a visual line scan camera is proposed in Figure 2, and multiple defect samples are collected to provide high-quality data for research on defect and impurity detection on the surface of rotors.
- An SAB-YOLOv5 surface defect detection algorithm is introduced, which incorporates the Atrous Spatial Pyramid Pooling (ASPP) structure and the Bidirectional Feature Pyramid Network (BiFPN) structure to enhance feature extraction and fusion. The algorithm also utilizes the Spatial Pyramid Dilated (SPD) convolutional structure to improve detection performance for minute defects on the surface.
- The improved model is tested and validated, and its performance is compared with that of other object detection model algorithms. The visual detection results are presented.

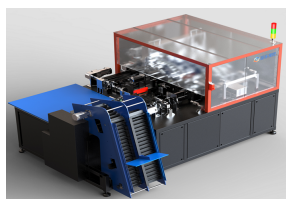


Figure 2. Experimental Rig based on SAB-YOLOv5.

The remainder of the paper is organized as follows: Section 2 outlines the constructed dataset. Section 3 delves into our methodological contributions in detail. Section 4 discusses the experimental environment and evaluation criteria. Section 5 presents comparative experiments and visualization results. Finally, Section 6 concludes the paper by summarizing the research findings and discussing prospects.

2. Data Acquisition

2.1. Images Acquisition

Currently, there is a scarcity of publicly available datasets for surface defect detection on the permanent magnetic ferrite magnet rotors. The dataset utilized in this study was collected on-site at the Zhengmin Magnetic Electric Group. The methodology involved employing a line-scan camera to scan the surface of magnetic rods and subsequently stitching the images to obtain complete surface representations of the cylindrical profile of the permanent magnetic ferrite magnet rotors like Figure 3. The authenticity of this dataset is rooted in its origin from practical field conditions as shown in Figure 2. This study categorizes and annotates four types of defects in Figure 4: edge breakage, crack, horizontal crack (crack-h), and hidden crack (subfissure). These four defect types represent the most commonly encountered flaws in the current target production process. The dataset comprises a total of 2240 images, annotated using the open-source tool LabelImg. The

dataset was divided into training, validation, and test sets in an 8:1:1 ratio as follows: 1792 images for training, 224 images for validation, and 224 images for testing.

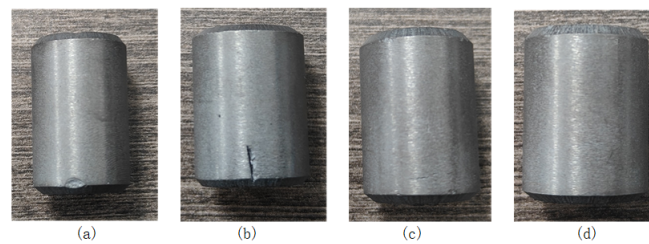


Figure 3. The collected the permanent magnetic ferrite magnet rotors samples: (a) edge breakage, (b) crack, (c) crack h, (d) subfissure.

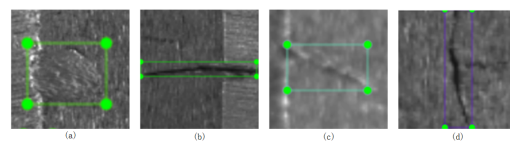


Figure 4. Data augmentations: (a) edge breakage, (b) crack, (c) crack h, (d) subfissure.

2.2. Image Annotation and Dataset Production

Annotation information is a vital component in the construction of for this study. The labeling tool was utilized to manually label the location and type information of various defects with intensive quality issues. Each labeled image generated a file which included the image name, image size, label name, and the location information of the labeled object. Given the image characteristics illustrated in Figure 4, the labels were divided into the following four categories: the final dataset was divided into a training set, a verification set, and a test set following a ratio to ensure no conflicts between the verification set and test set data. This division provided assurance for subsequent model accuracy evaluation.

To augment the model's generalization ability and mitigate overfitting, this study implemented data enhancement methods for the collected permanent magnetic ferrite magnet rotor image before training, as shown in Figure 4. By introducing different transformations such as rotation, scaling, translation, flipping, cropping, etc., data augmentation can help the model learn invariance under different transformations, thereby enhancing the model's robustness to various transformations and noise, and improving its generalization ability. Data augmentation can effectively reduce the overfitting of the model on the training set, because after seeing samples that underwent different transformations, the model is more likely to learn the essential features of the data rather than memorizing the details of the training set.

3. Methodologies

3.1. Overview of the Proposed Method

After analyzing the defect detection task on the surface of magnetic rods, this study identified several characteristics of defect data on the magnetic rod surface:

- (1) **Diverse Scale of Identical Features:** Features with the same characteristics exist at different scales, sometimes exhibiting significant differences. For example, features like edge collapse may vary in size from as small as 5×5 pixels to as large as 80×80 pixels.
- (2) **Detection of Small Targets:** Some defects are of exceedingly small size, presenting a challenge in the context of small object detection. Examples include fine crack structures.

The current YOLOv5 algorithm comprises a Convolutional Neural Network (CNN)-based visual feature extraction backbone and a detection head for predicting the category and bounding box of each object-containing region. To address the aforementioned characteristics, an additional Neck is introduced between the backbone and the detection head.

This Neck incorporates an extra Neck to combine features from multiple scales, generating semantically robust features suitable for detecting targets of varying sizes. Despite the use of an image pyramid structure to enhance the extraction of multi-scale features, the algorithm's performance is still inadequate in the context of small targets. Therefore, to optimize the multi-scale, small target issues present in the current dataset, three enhancements are introduced to the YOLOv5 algorithm as shown in Figure 5.

The modifications include introducing an SPD-Conv module [25] in the main network, replacing the original SPPF module [26] with an ASPP module [27], and augmenting the Neck section with a BiFPN-concat [28] structure. These modifications aim to strengthen the feature extraction capabilities for different scales and further enhance the network's feature fusion capabilities.

The ASPP (Atrous Spatial Pyramid Pooling) module and the BiFPN-concat (Bi-directional Feature Pyramid Network with concatenation) module are typically associated with feature fusion in semantic segmentation and object detection. In general, the ASPP module is commonly employed for semantic segmentation tasks, enhancing the model's understanding of image semantics by capturing context information at different scales. The BiFPN-concat module is typically used for object detection, facilitating feature map fusion at different levels and scales to improve detection performance.

The design of the SPD-Conv (Spatial Pyramid Dilated Convolution) module aims to enhance YOLOv5's detection performance on low-resolution images and small objects. This module introduces the concept of spatial pyramid dilated convolution, achieving feature perception at different scales by using convolutional kernels with varying dilation rates in the convolutional layer. This design effectively captures target information at different scales in the image, thereby improving the accuracy of small object detection.

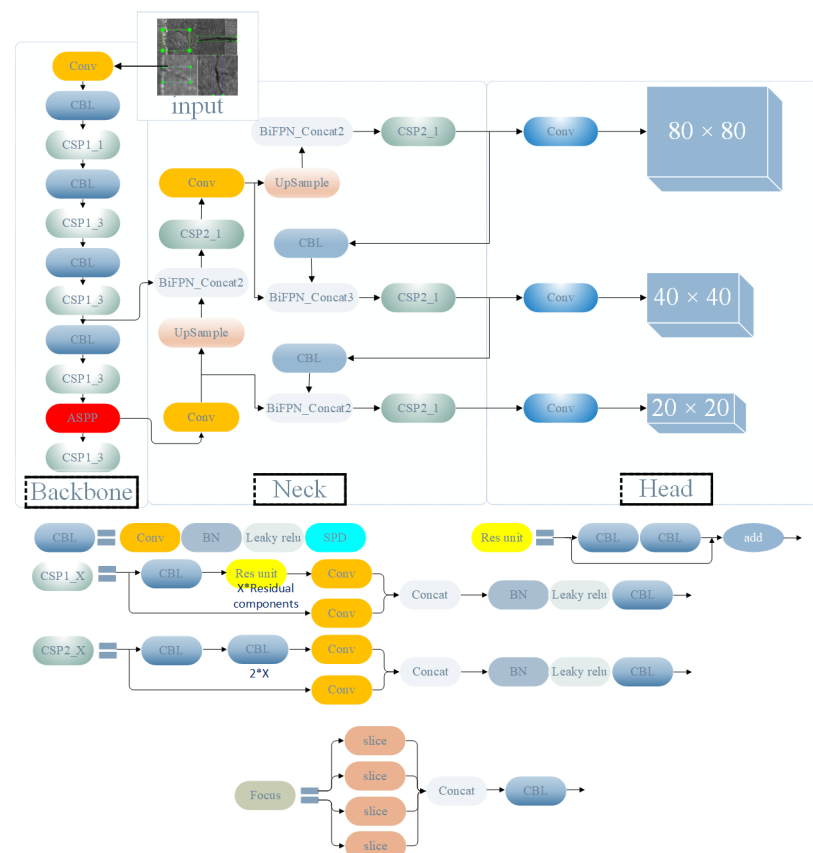


Figure 5. Modified YOLOv5x Architecture Diagram Structure.

3.2. About YOLO

Since the emergence of the YOLO (You Only Look Once) algorithm, it has consistently stood out as an advanced algorithm in the field of object detection. Particularly, YOLOv5 has been prioritized in various vision-related domains. YOLOv5 branches into YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x [29], which tailors its model scale to various sizes. Among them, the YOLOv5s model, characterized by its smaller size, exhibits parameters and computational efficiency that, when compared to the other three variants, are more suitable for real-time detection in industrial production, ensuring both speed and accuracy. The YOLOv5 architecture primarily comprises an input, a backbone, a Neck, and a head. It represents a one-stage anchor-free object detection algorithm, an evolution from its foundational predecessors.

The YOLOv5 model primarily comprises four components: the input, backbone feature extraction network, feature fusion network, and detection head. The Focus module, situated before the input, slices images to increase channel count, reducing parameters and computational load while enhancing recognition speed. The input involves preprocessing of images, and during the training phase, YOLOv5 employs Mosaic data augmentation to accelerate model training and improve network accuracy, mitigating memory requirements. Adaptive anchor box calculation enables backward updates during the training phase to adjust network parameters. Adaptive image scaling overcomes original image scaling issues, enhancing the overall algorithm's inference speed. The backbone network in YOLOv5 mainly includes Conv modules, C3 modules, and SPP modules. The Conv module consists of convolutional layers, BN layers, and activation functions. The convolutional layers extract local spatial information from input features, and BN layers normalize after convolution, expediting training and enhancing model generalization while reducing reliance on initialization. Activation functions introduce non-linear transformation capabilities to the neural network. The C3 module, also known as the CSP module, forms the basic unit of the CSPDarknet53 structure, increasing network depth and receptive field to enhance feature extraction. YOLOv5 introduces two CSP structures; in the YOLOv5s network, the CSP1-X structure applies to the backbone, increasing model depth and complexity to enhance detection accuracy. The SPP module is a pooling module that improves the recognition performance of the neural network. In YOLOv5, the Neck module employs the Feature Pyramid Network (FPN) pyramid structure. In object detection algorithms, the Neck module is typically used to combine feature maps from different hierarchical levels, generating feature maps with multiscale information to improve detection accuracy. In YOLOv5, the PANet feature fusion module is used as the Neck module. Another CSP2-X structure is applied to the feature fusion network, strengthening the network's feature fusion capability. The detection head utilizes multilevel feature fusion, primarily responsible for multiscale object detection on feature maps extracted by the backbone network. It fuses feature maps from different levels to obtain richer feature information, thereby improving the detection performance. This model demonstrates superior performance in detection.

3.3. Space to Depth

To address the problem of accelerating the identification or confirmation of defects, particularly in the context of the specific operation involving the magnetic rod (such as how Class is computed mathematically within the dataset), we introduce the SPD-conv, comprising a Spatial-to-Depth (SPD) layer and a non-strided convolutional layer. The subsequent discussion provides a detailed description of our SPD component, which extends a (primitive) image transformation technique for downsampling the feature maps within the CNN internally and across the entire CNN. We consider an intermediate feature map X of arbitrary size $S \times S \times C1$. The sub-feature map sequence is sliced as described below.

In general, for any (original) feature mapping X , sub-mapping $f(x, y)$ is composed of all feature mappings represented by feature maps $X(i, j)$, where $i + x$ and $j + y$ are divisible by a scaling factor. Consequently, each sub-map undergoes down-sampling of X by a scaling factor.

3.4. Adding the ASPP Module

To enhance the detection effectiveness of minute defects such as chipping edges and cracks on the surface of magnetic bars, this thesis introduces the SPD (Structural Preservation and Dilated) architecture into the backbone network. This integration aims to augment the network's capability for extracting features from small-scale targets.

The ASPP (Atrous Spatial Pyramid Pooling) module in Figure 6, a commonly used deep learning module for image semantic segmentation tasks, is introduced. The module aims to capture multi-scale contextual information to extract features more effectively, thereby enhancing the performance of the segmentation model. By employing different receptive field scales, the ASPP module facilitates a better understanding of objects and semantic information at various scales. The ASPP module essentially consists of a 1×1 convolution, a pooling pyramid, and ASPP pooling. The dilation factors for the layers of the pooling pyramid can be customized to achieve free multi-scale feature extraction.

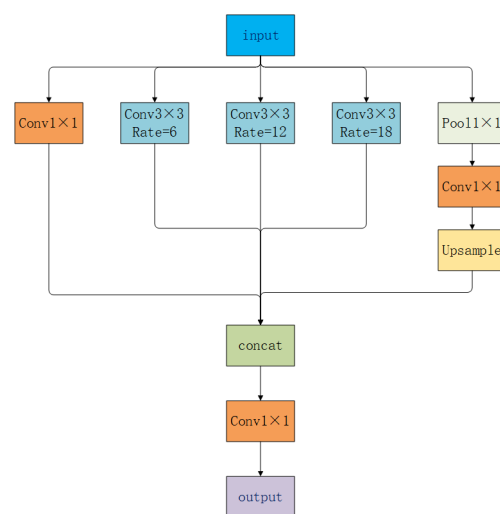


Figure 6. The structure of ASPP module.

- (1) **Atrous Convolution (Atrous Convolution):** The core of the ASPP module is the Atrous convolution, also known as dilated convolution. This convolutional operation allows for the convolutional kernel to skip a fixed number of pixel points on the input feature map, thus enlarging the receptive field. Atrous convolution aids in increasing the context information of features without introducing additional parameters.
- (2) **Multi-scale Sampling of Atrous Convolution:** The ASPP module typically employs multiple convolutional kernels with different dilation rates to obtain different receptive field scales. These kernels capture local details and broader contextual information, making the model more robust at different scales.
- (3) **Parallel Structure:** The ASPP module is often organized in a parallel manner, incorporating multiple Atrous convolution operations, each with a different dilation rate. The feature maps produced by these operations are concatenated along the channel dimension, forming a feature representation with multi-scale contextual information.
- (4) **Global Pooling:** To capture global contextual information, the ASPP module usually includes global average pooling, transforming the entire feature map into a single value. This value provides information about the entire image.
- (5) **Output:** The output of the ASPP module typically comprises multiple feature maps obtained after Atrous convolution and global pooling operations. These feature maps capture contextual information at different scales. They can be concatenated with other feature maps or a convolutional layer.

Due to the outstanding capability of ASPP to arbitrarily expand the receptive field without introducing additional parameters, the ASPP module is widely applied in target detection.

3.5. Utilization of the BiFPNConcat Layer

The original SPPF achieved the fusion of local and global features by performing multiple pooling operations on the original feature maps. In order to further enhance the capability for extracting multi-scale features, this study replaces the original SPPF module with the ASPP module in Figure 7. In the context of object detection tasks, a crucial aspect for successfully detecting key elements of objects lies in the effective capture and understanding of feature information across various scales and levels. This is attributed to the influence of multiple factors on the appearance and contextual information of objects, such as their size, orientation, posture, and degree of occlusion. A single object may manifest at different scales in an image, rendering feature extraction relying solely on a single scale inadequate for proficient object detection. The advent of feature fusion networks addresses this challenge. The primary objective of these networks is to seamlessly integrate feature maps from different convolutional layers and resolutions, thereby establishing a more comprehensive, multi-layered, and multi-scale feature representation. Through feature fusion, models can concurrently perceive both the global contextual information and detailed information of objects, irrespective of changes in object scale. This is crucial for the performance of object detection, which demands precise localization and classification of objects across diverse scales.

The utilization of feature fusion networks contributes to enhancing the detection accuracy of small- and large-scale objects while also addressing complexities such as object occlusion, varying angles, and postures. In summary, the significance of feature fusion networks in the field of object detection lies in their ability to facilitate a better understanding of multi-scale information, thereby improving detection performance and rendering models more robust to various real-world scenarios.

The concat layer is frequently employed to leverage semantic information from feature maps of different scales, enhancing performance by increasing channels. However, it is often advisable to perform concatenation after Batch Normalization (BN) to fully exploit its potential. Concatenation in the dimension is commonly used in multitask problems. Concatenation typically occurs unidirectionally, either from low-resolution features to high-resolution features or vice versa, without bidirectional information transfer.

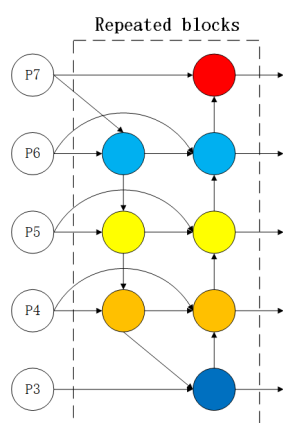


Figure 7. The structure of BiFPN module.

BiFPN is a network structure composed of different-scale feature pyramid levels. It enhances object detection algorithms by introducing multiple feature pyramid levels into the network and facilitating feature fusion through connecting feature maps between levels. This architecture improves the detection performance of small and distant objects, mitigating the unidirectional limitation of concat. For convenient and rapid multi-scale feature fusion, the BiFPNConcat module is employed before each CSP module in the Head section is expressed as in Figure 7.

4. Implementation Details

4.1. Experimental Rig

To automate the acquisition and processing of images of cylindrical magnet surfaces, we designed an experimental device as shown in Figure 8. This device is a machine vision system and this study describes the components of a packaging device. Here is a brief explanation of each component:

- (1) Frame: The supporting structure of the entire device.
- (2) Feeding lifting mechanism: A mechanism used to lift materials from the bottom to the top, including the chain plate conveyor lifting mechanism and the material distribution baffle. The chain plate conveyor lifting mechanism is fixed on the frame and has two channels, and the material distribution baffle is used to distribute the materials into the two channels.
- (3) Feeding conveyor mechanism: A mechanism used to transport materials from the feeding lifting mechanism to the next step. It includes the first conveyor belt, the first partition baffle, the preliminary screening detection component, the preliminary screening waste kicking component, and the material pushing component. The first conveyor belt is installed on the frame, and the materials are transported from the chain plate conveyor lifting mechanism to the first conveyor belt. The first partition baffle is set along the length of the first conveyor belt, dividing it into a middle channel and two side material channels. The preliminary screening detection component is used to detect half-cut materials or materials with large-sized defects, and the preliminary screening waste kicking component is placed in the middle channel to kick the waste materials into the waste kicking channel. The material pushing component is located at the end of the material channels.
- (4) Main detection mechanism: Fixed on the frame, it includes the supporting star wheel, the detection camera, and the material discharging component. The supporting star wheel has multiple slots to catch the materials. Two supporting star wheels arranged in parallel correspond to the position of the material pushing component, and the material discharging component is located on the side of the supporting star wheel away from the material pushing component. The detection cameras are evenly distributed on the top of the supporting star wheel.
- (5) Material discharging conveyor mechanism: Fixed on the frame, located below the material discharging component, used to transport materials out of the device.
- (6) Boxing mechanism: Fixed on the frame, it is located below the material discharging conveyor mechanism and used for boxing the materials.

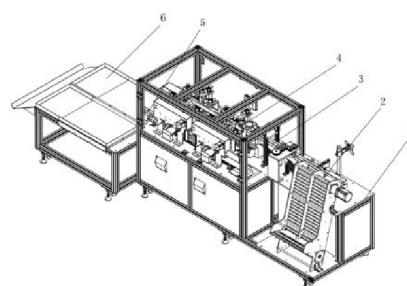


Figure 8. The structure of Experimental Rig. (1) Frame, (2) Feeding lifting mechanism, (3) Feeding conveyor mechanism, (4) Main detection mechanism, (5) Material discharging conveyor mechanism, (6) Boxing mechanism.

When the device is activated, the materials are lifted from the storage area to the inspection area using an elevation mechanism. In the inspection area, the conveyor belt separates the cylindrical magnets at regular intervals to facilitate further inspection and processing. Once the magnets are separated, the visible light camera captures images of the surface of each magnet. The circular light source provides uniform illumination to

ensure good image quality. The captured images are then transmitted to the computer for processing. The computer processes the received images to extract the features of the magnet surfaces. This may involve algorithms such as edge detection, color analysis, and shape matching. The processed images can be used for subsequent analysis and identification. Based on the features of the magnet surfaces, the computer determines the type or characteristics of each magnet and classifies them accordingly. The classified magnets are then placed into their corresponding boxes for storage and future use. With this experimental device, the automatic acquisition and processing of images of cylindrical magnet surfaces can be achieved. This plays a significant role in the automation and efficiency improvement of magnet production and quality control processes.

4.2. Experimental Environment

In this study, all experiments were conducted based on the Pytorch deep learning framework with the programming language of Python. The hardware is configured with an Intel(R)Core(TM) i9-10980XE CPU @ 3.00GHz, 128 GB of RAM, NVIDIA GeForce RTX 3090(24 GB) GPU, and a Ubuntu 18.04 operating system. The main training parameters are listed in Table 1.

Table 1. Hardware and software environment for model training.

Item	Type
CPU	Intel(R) Core(TM) i9-10980XE CPU @ 3.00 GHZ
Memory	128 G
GPU	NVIDIA GeForce RTX 3090 24 G *2
Operating System	Ubuntu 18.04.6 LTS(64-bit)
CUDA Version	11.3
Deep Learning Frame	Pytorch

To assess the viability of the proposed methodology, an incremental analysis of enhancement points was conducted through ablation experiments. Subsequently, the algorithm's performance underwent meticulous scrutiny via ablation experiments, followed by a comparative assessment against a widely accepted mainstream algorithm. The primary evaluation metrics utilized for assessing the algorithm's detection performance included mean average precision (mAP), frames per second (FPS), and weight size measured in megabytes (MBs).

4.3. Evaluation Criteria

To assess the performance of the Yolov5 combination method, several evaluation metrics are employed, including precision (P), recall (R), mean average precision (mAP), average precision (AP), floating-point operations per second ($FLOPs$), and parameter count. P and R are precisely defined in the equations, where TP represents the number of positive samples predicted as positive, FN represents the count of negative samples incorrectly predicted as positive, and FP denotes the count of negative samples incorrectly predicted as positive. AP denotes the average accuracy of each category in the detection, mAP stands for mean average precision of multiple categories in object detection, P represents the proportion of correctly predicted positive samples out of all positive predictions made by the detection model, whereas R indicates the proportion of correctly predicted positive samples out of all actual positive samples. P , R , AP , and mAP calculation formulae are as follows:

$$P = \frac{TP}{TP + FP} \times 100\% \quad (1)$$

$$R = \frac{TP}{TP + FN} \times 100\% \quad (2)$$

$$AP = \int_0^1 P(R) dR \quad (3)$$

$$mAP = \frac{\sum_{i=1}^n AP(i)}{n} \% \quad (4)$$

In this section, we address the interconnection of parameters and their influence on defect detection of small targets on the surface of the permanent magnetic ferrite magnet rotor. Building upon the solutions provided by the preceding formulas for defect detection on the surface of the permanent magnetic ferrite magnet rotors and the subsequent categorization of flaws into classes, we delve into the issues concerning the impact of specific images on accuracy. In the realm of published research, the YOLOv5 series presents a range of models with varying trade-offs between speed and detection performance. Starting with YOLOv5s Figure 9, the smallest version boasts the fastest speed but comes at the cost of the lowest detection performance. Moving on to YOLOv5m, a medium-sized model sacrifices a bit of speed compared to YOLOv5s yet showcases improved detection capabilities. Progressing further, YOLOv5n offers a speed boost over YOLOv5m while further enhancing detection performance. Stepping up in size, the YOLOv5l model slows down compared to YOLOv5n but delivers superior detection performance. Finally, YOLOv5x, the largest version in the series, may be the slowest in speed, yet shines with the best detection performance among its counterparts.

Model	size (pixels)	mAP ^{val} 0.5:0.95	mAP ^{val} 0.5	Speed CPU b1 (ms)	Speed V100 b1 (ms)	Speed V100 b32 (ms)	params (M)	FLOPs @640 (B)
YOLOv5n	640	28.0	45.7	45	6.3	0.6	1.9	4.5
YOLOv5s	640	37.4	56.8	98	6.4	0.9	7.2	16.5
YOLOv5m	640	45.4	64.1	224	8.2	1.7	21.2	49.0
YOLOv5l	640	49.0	67.3	430	10.1	2.7	46.5	109.1
YOLOv5x	640	50.7	68.9	766	12.1	4.8	86.7	205.7

Figure 9. Parameter situations under different pre-trained models

The choice of SAB-YOLOv5X is based on several reasons Table 2. With a parameter count of 294,308,409, this model falls between other models in terms of parameter size, striking a balance without being overly large. It also possesses 429.6 GFLOPs of computational power, higher than YOLOv5X and YOLOv5 adding the ASPP structure but still manageable. GPU memory consumption stands at 4.316, slightly higher than that of some models but within acceptable limits. In terms of inference speed, it has a forward time of 47.32 milliseconds and a backward time of 79.05 milliseconds, both within reasonable ranges. The advantages of SAB-Yolov5X lie in its balanced parameter count and computational power, making it suitable for environments with limited computing resources. Despite not being the fastest, its inference speed remains relatively quick given the balance in parameter count and computational power. The model exhibits strong feature fusion capabilities with BiFPN effectively merging features at different levels to enhance object detection across various scales. The ASPP module captures multi-scale information through different dilation rates, improving pixel-level prediction accuracy. The SPD method facilitates effective knowledge distillation from complex models to simplified ones, enhancing model performance and inference speed. By integrating multiple feature

extraction techniques, this model performs well in object detection tasks. Furthermore, its relatively lower memory consumption of 4.316 GPU memory remains acceptable in current hardware environments. Therefore, the improved model aims to strike a balance between parameter count, computational power, inference speed, and memory usage. The model combines various effective feature extraction techniques, resulting in strong performance capabilities for object detection tasks.

Table 2. Comparison of data with different additional modules

Model	Para	GFLOPs	GPU Mem	Forward	Backward
YOLOv5X	292695610	204.7	2.152	38.26	50.96
YOLOv5X+ASPP	95838001	364.6	3.93	46.28	79.88
YOLOv5X+SPD	87674170	313	2.814	35.47	57.64
YOLOv5X+BIFPN	306427450	219.2	2.288	36	60.11
YOLOv5X+ASPP+BIFPN	306427450	388.9	4.156	48.49	85.67
YOLOv5X+ASPP+SPD	323108081	490.3	4.482	53.22	82.10
SAB-YOLOv5	294308409	429.6	4.316	47.32	79.05

We delineate various categories of images, defining several problems associated with them. Through the application of mathematical formulas, we resolve these issues, thereby enhancing the precision of our defect detection methodology. After training, a comparative test is conducted using the test dataset. In total, 224 images to be detected in the test dataset are input into the trained model for testing, and the results are shown in Table 3. As can be seen from the comparison results, the *mAP* value using YOLOv5+SPD is 86.1%, 1.9% over the original YOLOv5X, indicating the SPD can act as an optimized clustering role, improve the accuracy of the detection algorithm, and strengthen localization. The *mAP* of the ASPP model is 3.0% higher than that of the YOLOv5X model, and the BIFPN model is improved by 5.0%. Our SAB-YOLOv5X data augmentation method improved *mAP* by 14.1% compared to the original YOLOv5X model.

Table 3. The impact of attention mechanism module on model performance.

Methods	mAP (%)	Precision (%)	Recall (%)	F1-Score
YOLOv5X	84.2	84.0	78.7	80.3
YOLOv5X+SPD	86.1	85.5	81.4	83.4
YOLOv5X+ASPP	87.2	82.8	84.2	83.5
YOLOv5X+BIFPN	89.2	87.0	84.6	85.8
SAB-YOLOv5	98.3	97.1	96.0	98.3

Simultaneously, our approach allows for the recognition of defects at different scales within the same timeframe. The optimization of the method is substantiated by contrasting the data obtained through defect detection under various conditions. Moreover, we expound upon the interpretation of data resulting from diverse datasets and computational processes. This comprehensive analysis serves to validate the efficacy of the proposed method in addressing the challenges posed by defect detection on the surface of the permanent magnetic ferrite magnet rotors.

5. Experimental Results and Analysis

5.1. Ablation Experiments

In order to explore the impact of different structures on the results and validate the effectiveness of the improved algorithm for detecting magnetic rod defects, the following ablation experiments were conducted in the present study. In these experiments, YOLOv5 was used as the baseline algorithm, and the mean Average Precision *mAP* metric was employed to evaluate performance. Various improvements were applied to the original algorithm, and the results are summarized in Figure 10. The comparative results indicate

that, when the SPD module is incorporated into the baseline algorithm, the algorithm's mAP for enhanced feature extraction on small targets is 86.1%, which is 1.9% higher than the mAP using the original YOLOv5. Utilizing the ASPP module with different receptive field scales further enhances the algorithm's mAP for multi-scale feature extraction to 87.2%, achieving a 3% increase in average precision compared to the original experiment. Substituting the BiFPN module for the original YOLOv5 feature fusion network improves multi-scale target detection and testing accuracy, resulting in an accuracy of 89.6%, which is 5.4% higher than when using the original YOLOv5 mAP . The complete improved algorithm, when compared to the original YOLOv5 algorithm, demonstrates a 14.1% increase in mAP . Overall, the experimental results indicate that the optimization of each algorithm has a positive impact on the final results. The modular algorithm construction allows for detection accuracy, further enhancing overall model performance.

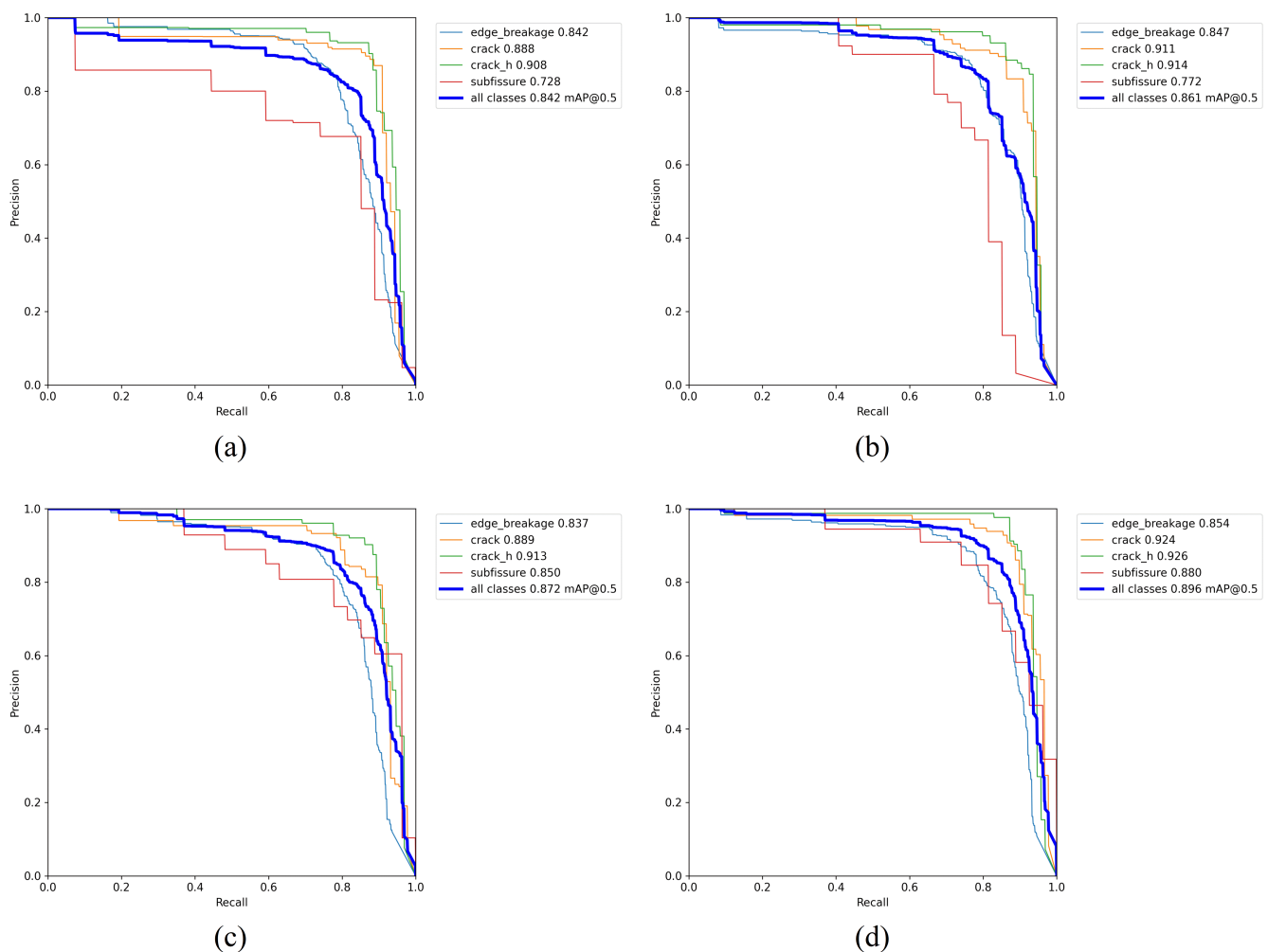


Figure 10. Attention Mechanism including, respectively, (a) YOLOv5, (b) YOLOv5+SPD, (c) YOLOv5+ASPP, and (d) YOLOv5+BiFPN

To verify the feasibility of different models on the Zhengmin datasets, we conducted training and testing on several mainstream one-stage and two-stage network models. We comprehensively considered precision P , recall R , mean Average Precision mAP , model parameters (Params), model size, model computational costs (BFLOPs), and detection speed as comparative indicators. The experimental results are presented in the following Table 4, with YOLOv5s, YOLOv5m, YOLOv5x, Faster R-CNN, YOLOv8, and the improved YOLOv5s selected for comparison experiments. Faster R-CNN is the most typical two-stage object detection model with the highest number of parameters and computational cost

in the comparative experiments. Its mAP is higher than those of YOLOv5s, YOLOv5m, YOLOv5x, and YOLOv8. However, due to its less-than-ideal performance in detecting small objects, the mAP is not very high, and both parameter count (Params) and computational cost are also high.

Table 4. Comparison of experimental results of each model.

Methods	mAP (%)
Yolov5s	75.6
Yolov5m	82.1
Yolov5x	84.2
Faster RCNN	88.6
Yolov8x	88.1
Ours	98.3

5.2. Visualization and Discussion

Figure 10 above illustrates the AP (Average Precision) comparison between the YOLOv5 algorithm and the modified algorithm across various defect categories. From the chart, it is evident that the proposed enhancement strategies effectively enhance the detection accuracy of surface defects on permanent magnetic ferrite magnet rotors. There is a notable improvement in the detection of various defect categories. For challenging cases such as edge-breakage and sub fissure, the modified model outperforms the original network significantly, validating the improved model's capability to achieve notable enhancements in handling difficult samples.

5.3. Comparison Results on Different Models

To verify the feasibility of different models on the Zhengmin datasets, we conducted training and testing on several mainstream one-stage and two-stage network models. We comprehensively considered precision (P), recall (R), mean Average Precision (mAP), model parameters (Params), model size, model computational costs (BFLOPs), and detection speed as comparative indicators. The experimental results are presented in the following table, with YOLOv5s, YOLOv5m, YOLOv5x, Faster R-CNN [29], YOLOv8, and the improved YOLOv5s selected for comparison experiments. Faster R-CNN is the most typical two-stage object detection model with the highest number of parameters and computational cost in the comparative experiments. Its mAP is higher than those of YOLOv5s, YOLOv5m, YOLOv5x, and YOLOv8. However, due to its less-than-ideal performance in detecting small objects, the mAP is not very high, and both parameter count (Params) and computational cost are also high.

Through the comparative analysis of the data presented in Figure 10, it is evident that the improved YOLOv5x algorithm significantly increased the mean Average Precision (mAP) from 84.2% to 98.3%, marking a remarkable improvement of 14.1 percentage points. Specifically, the Average Precision (AP) for the “edge-breakage” class rose from 84.2% to 97.8%, demonstrating a gain of 13.6 percentage points. The “crack” class AP increased from 88.8% to 99.3%, showing an improvement of 10.5 percentage points. Similarly, the AP values for “crack-h” rose from 90.8% to 97.9%, representing a gain of 7.1 percentage points. Notably, the AP for “point-etch-surface” increased from 72.8% to 98.1%, indicating a boost of 6.9 percentage points. The detection performance for all four defect types of experienced improvement. Figure 11 shows the visualization results. SAB-YOLO5X demonstrates exceptional proficiency in accurately predicting the positions of defects on surfaces and targets of varying sizes within images. The refined model not only excels in identifying challenging surface imperfections, but also exhibits precision in classifying and fulfilling the requirements of precise object detection.

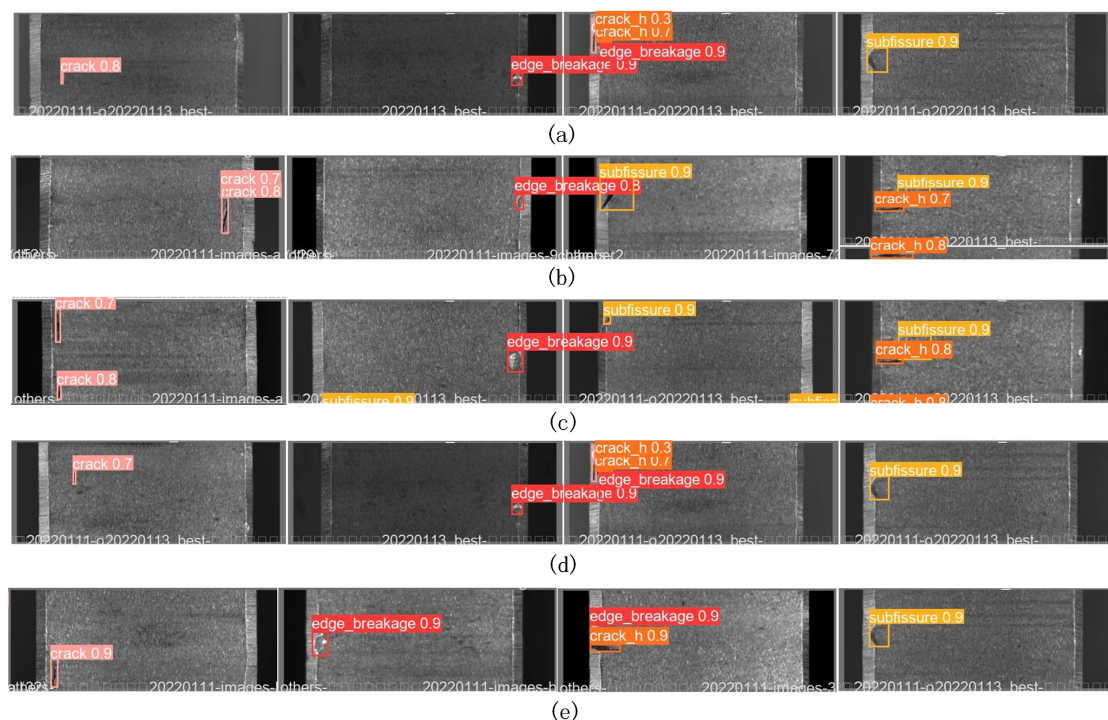


Figure 11. Visualization of detection results on the permanent magnetic ferrite magnet rotors on YOLOv5 models including, respectively, (a) YOLOv5, (b) YOLOv5+SPD, (c) YOLOv5+ASPP, (d) YOLOv5+BiFPN and (e) SAB-YOLOv5.

6. Conclusions

This study observes that various defects on the surface of permanent magnetic ferrite magnet rotors, such as edge breakage, cracks, and hidden cracks, exhibit similar features but vary significantly in scale. To address this, the ASPP module is introduced, incorporating depth-wise separable convolutions and dilated convolutions with different rates. This design replaces the pooling operation in the original YOLOv5 Spatial Pyramid Pooling (SPP) with a new feature pyramid model that combines global average pooling. This facilitates the aggregation of multi-scale contextual information, enhancing the model's ability to recognize objects of the same type but different sizes. Additionally, the introduction of the BiFPN structure in the aggregation network improves feature fusion, further enhancing the algorithm's ability to identify defects of different scales. As surface images are captured by rotating the magnetic rod through a complete revolution, creating a stitched image from the motion-captured surface images using a line scan camera inevitably introduces some blurry images. To address this, the SPD convolution module is introduced at the output end, proving beneficial for improving the detection accuracy of fuzzy images and effectively identifying smaller defect targets. Experimental results validate the effectiveness of the enhanced algorithm, demonstrating improved defect detection performance on common target surface defects. However, it is acknowledged that the current dataset is relatively limited in terms of diversity. Future work will focus on expanding the dataset to include a broader range of defect types.

Author Contributions: Conceptualization, B.Y., Q.L. and W.J.; Methodology, B.Y. and W.J.; Formal analysis, B.Y.; Investigation, Y.Z.; Resources, W.J.; Data curation, B.Y., S.Z. and Y.Z.; Writing—original draft, B.Y. and W.J.; Writing—review & editing, Q.L. and S.Z.; Project administration, W.J. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data are contained within the article.

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

References

1. Egorov, D.; Petrov, I.; Pyrhönen, J.; Link, J.; Stern, R. Hysteresis loss in ferrite permanent magnets in rotating electrical machinery. *IEEE Trans. Ind. Electron.* **2018**, *65*, 9280–9290. [\[CrossRef\]](#)
2. Lauwers, B.; Klocke, F.; Klink, A.; Tekkaya, A.E.; Neugebauer, R.; McIntosh, D. Hybrid processes in manufacturing. *CIRP Ann.* **2014**, *63*, 561–583. [\[CrossRef\]](#)
3. De Gennaro, M.; Jürgens, J.; Zanon, A.; Gragger, J.; Schlemmer, E.; Fricassè, A.; Marengo, L.; Ponick, B.; Olabarri, E.T.; Kinder, J.; et al. Designing, prototyping and testing of a ferrite permanent magnet assisted synchronous reluctance machine for hybrid and electric vehicles applications. *Sustain. Energy Technol. Assess.* **2019**, *31*, 86–101. [\[CrossRef\]](#)
4. Tao, J.; Wang, Y.; Wang, K. Defect detection for end surface of ferrite magnetic tile. In Proceedings of the 8th International Symposium on Advanced Optical Manufacturing and Testing Technologies: Optical Test, Measurement Technology, and Equipment, Suzhou, China, 26–29 April 2016; SPIE: Bellingham, WA, USA, 2016; Volume 9684, pp. 513–520.
5. Gallimore, E.; Terrill, E.; Pietruszka, A.; Gee, J.; Nager, A.; Hess, R. Magnetic survey and autonomous target reacquisition with a scalar magnetometer on a small AUV. *J. Field Robot.* **2020**, *37*, 1246–1266. [\[CrossRef\]](#)
6. Li, M.; Jia, J.; Lu, X.; Zhang, Y. A method of surface defect detection of irregular industrial products based on machine vision. *Wirel. Commun. Mob. Comput.* **2021**, *2021*, 6630802. [\[CrossRef\]](#)
7. Sultana, F.; Sufian, A.; Dutta, P. Advancements in image classification using convolutional neural network. In Proceedings of the 2018 Fourth International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN), Kolkata, India, 22–23 November 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 122–129.
8. Kim, J.; Um, S.; Min, D. Fast 2D complex Gabor filter with kernel decomposition. *IEEE Trans. Image Process.* **2017**, *27*, 1713–1722. [\[CrossRef\]](#) [\[PubMed\]](#)
9. Aminzadeh, M.; Kurfess, T. Automatic thresholding for defect detection by background histogram mode extents. *J. Manuf. Syst.* **2015**, *37*, 83–92. [\[CrossRef\]](#)
10. Chen, X.; Fu, C.; Tie, M.; Sham, C.W.; Ma, H. AFFNet: An Attention-Based Feature-Fused Network for Surface Defect Segmentation. *Appl. Sci.* **2023**, *13*, 6428. [\[CrossRef\]](#)
11. Liu, J.; Shen, X.; Wang, J.; Jiang, L.; Zhang, H. An intelligent defect detection approach based on cascade attention network under complex magnetic flux leakage signals. *IEEE Trans. Ind. Electron.* **2022**, *70*, 7417–7427. [\[CrossRef\]](#)
12. 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, 27–30 June 2016; pp. 779–788.
13. Yue, X.; Wang, Q.; He, L.; Li, Y.; Tang, D. Research on tiny target detection technology of fabric defects based on improved Yolo. *Appl. Sci.* **2022**, *12*, 6823. [\[CrossRef\]](#)
14. Zuo, Y.; Wang, J.; Song, J. Application of YOLO object detection network in weld surface defect detection. In Proceedings of the 2021 IEEE 11th Annual International Conference on CYBER Technology in Automation, Control, and Intelligent Systems (CYBER), Jiaxing, China, 27–31 July 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 704–710.
15. Hussain, M. YOLO-v1 to YOLO-v8, the Rise of YOLO and Its Complementary Nature toward Digital Manufacturing and Industrial Defect Detection. *Machines* **2023**, *11*, 677. [\[CrossRef\]](#)
16. Zhong, Y.; Wang, J.; Peng, J.; Zhang, L. Anchor box optimization for object detection. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Snowmass, CO, USA, 1–5 March 2020; pp. 1286–1294.
17. Murthy, C.B.; Hashmi, M.F.; Muhammad, G.; AlQahtani, S.A. YOLOv2PD: An efficient pedestrian detection algorithm using improved YOLOv2 model. *Comput. Mater. Contin.* **2021**, *69*, 3015–3031.
18. Zhang, X.; Gao, Y.; Wang, H.; Wang, Q. Improve YOLOv3 using dilated spatial pyramid module for multi-scale object detection. *Int. J. Adv. Robot. Syst.* **2020**, *17*, 1729881420936062. [\[CrossRef\]](#)
19. He, K.; Zhang, X.; Ren, S.; Sun, J. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *37*, 1904–1916. [\[CrossRef\]](#) [\[PubMed\]](#)
20. Zhang, J.; Zhang, H.; Liu, B.; Qu, G.; Wang, F.; Zhang, H.; Shi, X. Small object intelligent detection method based on adaptive recursive feature pyramid. *Heliyon* **2023**, *9*, e17730. [\[CrossRef\]](#) [\[PubMed\]](#)
21. Liao, X.; Lv, S.; Li, D.; Luo, Y.; Zhu, Z.; Jiang, C. YOLOv4-MN3 for PCB surface defect detection. *Appl. Sci.* **2021**, *11*, 11701. [\[CrossRef\]](#)
22. Liu, R.; Chen, Y.; Wang, J.; Guo, Z. Attentive Mix: An Efficient Data Augmentation Method for Object Detection. In Proceedings of the 2021 7th International Conference on Computer and Communications (ICCC), Chengdu, China, 10–13 December 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 770–774.
23. Gao, X.; Yu, A.; Tan, J.; Gao, X.; Zeng, X.; Wu, C. GSD-YOLOX: Lightweight and more accurate object detection models. *J. Vis. Commun. Image Represent.* **2024**, *98*, 104009. [\[CrossRef\]](#)
24. Yu, G.; Zhou, X. An Improved YOLOv5 Crack Detection Method Combined with a Bottleneck Transformer. *Mathematics* **2023**, *11*, 2377. [\[CrossRef\]](#)
25. Li, A.; Sun, S.; Zhang, Z.; Feng, M.; Wu, C.; Li, W. A Multi-Scale Traffic Object Detection Algorithm for Road Scenes Based on Improved YOLOv5. *Electronics* **2023**, *12*, 878. [\[CrossRef\]](#)
26. Liu, H.; Sun, F.; Gu, J.; Deng, L. Sf-yolov5: A lightweight small object detection algorithm based on improved feature fusion mode. *Sensors* **2022**, *22*, 5817. [\[CrossRef\]](#)

27. Ji, Z.; Wu, Y.; Zeng, X.; An, Y.; Zhao, L.; Wang, Z.; Ganchev, I. Lung nodule detection in medical images based on improved YOLOv5s. *IEEE Access* **2023**, *11*, 76371–76387. [[CrossRef](#)]
28. Ye, M.; Qu, X.; Huang, J.; Wu, X. In-air Handwriting System Based on Improved YOLOv5 algorithm and Monocular Camera. In Proceedings of the 2022 IEEE 34th International Conference on Tools with Artificial Intelligence (ICTAI), Macao, China, 31 October–2 November 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 949–954.
29. Jocher, G.; Stoken, A.; Borovec, J.; Changyu, L.; Hogan, A.; Diaconu, L.; Poznanski, J.; Yu, L.; Rai, P.; Ferriday, R.; et al. *ultralytics/yolov5*, V3.0; Zenodo: Geneva, Switzerland, 2020.

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.