

## Article

# Real-Time Steel Surface Defect Detection with Improved Multi-Scale YOLO-v5

Ling Wang <sup>1,2</sup>, Xinbo Liu <sup>3,\*</sup>, Juntao Ma <sup>4</sup>, Wenzhi Su <sup>4</sup> and Han Li <sup>5</sup>

- <sup>1</sup> College of Chemistry and Materials Engineering, Hainan Vocational University of Science and Technology, Haikou 571156, China
- <sup>2</sup> Liaoning Key Laboratory of Chemical Additive Synthesis and Separation, Yingkou Institute of Technology, Yingkou 115014, China
- <sup>3</sup> SolBridge International School of Business, Woosong University, Daejeon 34613, Republic of Korea
- <sup>4</sup> Fulin Warehousing Logistics (Yingkou) Co., Ltd., Yingkou 115007, China
- <sup>5</sup> School of Electronics and Information Engineering, Liaoning University of Technology, Jinzhou 121001, China
- \* Correspondence: lxb715@hotmail.com

**Abstract:** Steel surface defect detection is an important issue when producing high-quality steel materials. Traditional defect detection methods are time-consuming and uneconomical and require manually designed prior information or extra supervisors. Surface defects have different representations and features at different scales, which make it challenging to automatically detect the locations and defect types. This paper proposes a real-time steel surface defect detection technology based on the YOLO-v5 detection network. In order to effectively explore the multi-scale information of the surface defect, a multi-scale explore block is especially developed in the detection network to improve the detection performance. Furthermore, the spatial attention mechanism is also developed to focus more on the defect information. Experimental results show that the proposed network can accurately detect steel surface defects with approximately 72% mAP and satisfies the real-time speed requirement.

**Keywords:** steel surface defect detection; deep learning; convolutional neural network



**Citation:** Wang, L.; Liu, X.; Ma, J.; Su, W.; Li, H. Real-Time Steel Surface Defect Detection with Improved Multi-Scale YOLO-v5. *Processes* **2023**, *11*, 1357. <https://doi.org/10.3390/pr11051357>

Academic Editors: Adel Ali Ahmed, AbdulRahman Alsewari, Yousef Fazea and Waleed Ali

Received: 27 February 2023

Revised: 9 April 2023

Accepted: 25 April 2023

Published: 28 April 2023



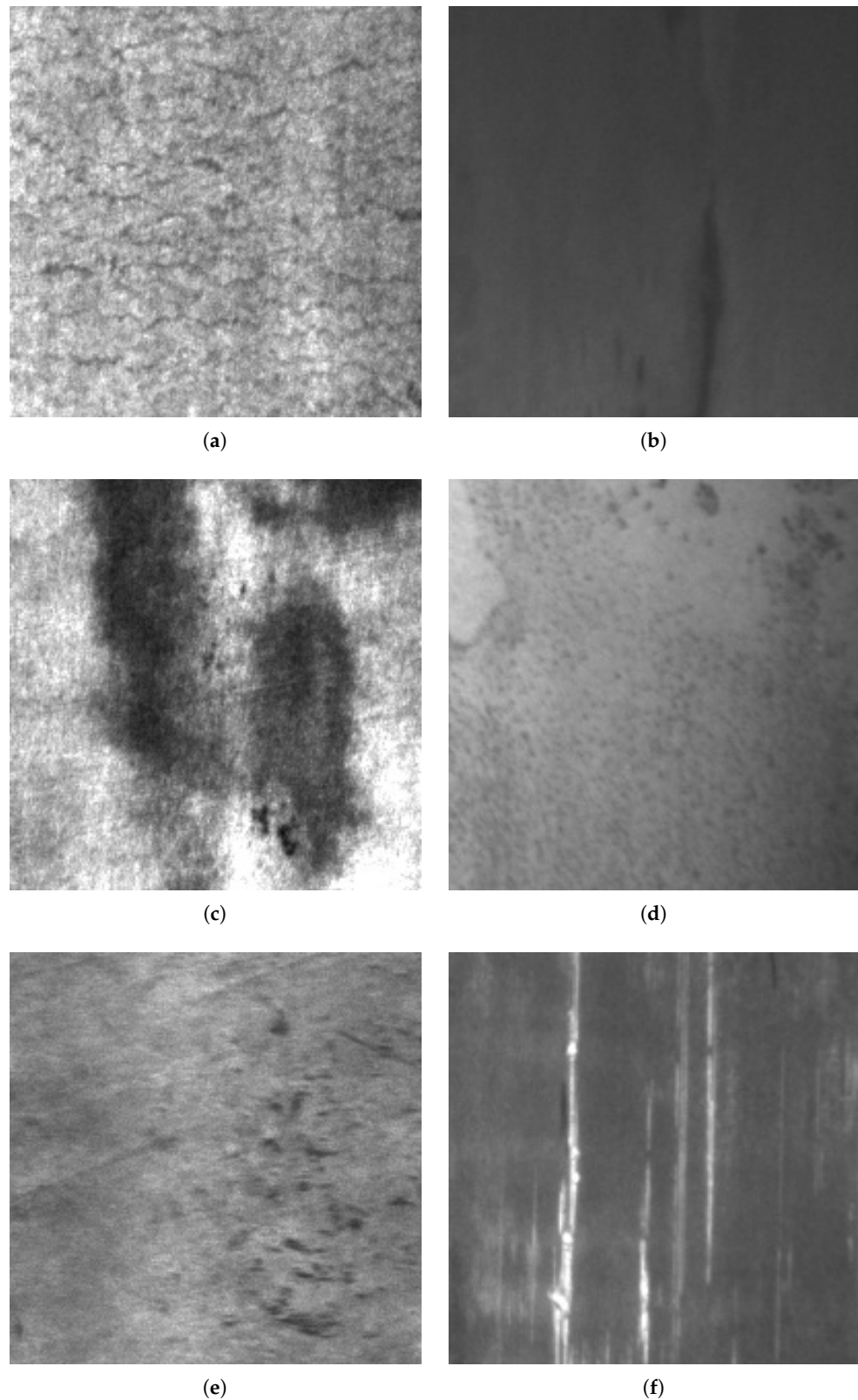
**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Steel surface defect detection is an important topic in material science research [1]. As one of the most important fundamental materials, steel contributes to numerous industry productions, such as airplanes, automobiles and high-speed railways. Among various steel productions, flat steel is the dominant product and contributes the most to industrial applications. As such, the quality of flat steel is vital for daily life.

Unfortunately, there are usually defects on flat steel surfaces, making it challenging to generate high-quality steel industrial productions. There are six typical defects on steel surfaces: crazing, inclusion, patches, pitted surface, rolled-in scale and scratches [2]. Figure 1 demonstrates the typical types of different steel surface defects in the North East University Detection (NEU-DET) dataset. The defects lead to bad quality in the flat steel, making it challenging to produce high-quality industrial productions.

In traditional factories, steel surface defect detection relies on human supervision, which is time-consuming and uneconomical [3]. On the one hand, extra supervisors require more resources than automatic detection does. On the other hand, human-dependent detection cannot ensure quality supervision around-the-clock. With the development of the industrial vision area, computers come to be a powerful tool to detect surface defects. Previous works rely on hand-crafted feature extractors and machine learning-based classifiers to localize the defects. An artificial neural network (ANN), support vector machine (SVM), k-nearest neighbor (KNN) and other machine learning technologies have been widely applied in different steel surface defect detection methods [4–6]. However, these works suffer from lower precision and cannot satisfy the real-time speed requirement.



**Figure 1.** Typical types of steel surface defects in the NEU-DET dataset: (a) Craze. (b) Inclusion. (c) Patches. (d) Pitted Surface. (e) Rolled-in Scale. (f) Scratches.

To boost the accuracy and improve the speed, there are convolutional neural networks (CNNs) especially developed for defect detection. Mustafa et al. used different image classification methods to recognize the diverse steel surface defects [7]. He et al.

utilized a multi-level feature fusion network and classified the different kinds of defects [8]. These works demonstrate good performances with two-stage object detection networks, separating the localization and classification steps, which is time-consuming.

In contrast to two-stage networks, you-only-look-once (YOLO) series methods utilize one-stage objection detection technology and achieve real-time speed [9–13]. In particular, the fifth version, YOLO-v5, achieves the state-of-the-art detection performance which has been widely utilized in various situations, such as letter recognition [13], circuit defect detection [14] and fabric detection [15]. However, the traditional YOLO-v5 method cannot effectively explore steel surface defects. On the one hand, as shown in Figure 1, there are different representations of the defects, making it difficult to accurately localize and classify the defect areas. On the other hand, the defects vary from different scales. The extra small or large defects are challenging to explore and detect.

This paper proposes an improved multi-scale YOLO-v5 technology for real-time steel surface defect detection. In particular, we develop a multi-scale block to effectively explore the defects. Convolutions with different filter sizes are especially developed to process the input images and generate the multi-scale information. The multi-scale image features are aggregated by one convolutional layer for information fusion, which boosts the representation capacity of the network. Furthermore, a spatial-attention mechanism is developed to concentrate more on the defect areas and improve the detection accuracy. The experimental results show that the improved multi-scale YOLO-v5 method can more accurately detect the steel surface defects than the original version, which satisfies the real-time speed requirement.

Our contributions can be concluded as follows:

- We propose an improved multi-scale YOLO-v5 network for effective steel surface defect detection, which achieves a high detection accuracy and demonstrates a good robust performance.
- We develop the multi-scale block and spatial attention mechanism to process the steel surface images, which effectively explore the defect information and improve the accuracy of the network.
- Experimental results show that the improved network has a higher prediction accuracy than the vanilla YOLO-v5 method, which satisfies the real-time speed requirement.

## 2. Related Work

### 2.1. Steel Surface Defect Detection

Steel surface defect detection is one of the most important tasks in the industrial vision research area. The task of steel surface defect detection is to automatically find the defects of the steel and guarantee the quality of industrial productions. In previous works, different computer vision and machine learning-based methods have been used to accurately detect defects. There are researchers utilizing the probability model to describe the steel surfaces without defects, and regarding the outliers as the examples with bad quality [16]. To this end, a dynamic threshold technology is developed to detect the outliers. Wang et al. used the histogram of image features to model the difference between the defect-free examples and bad cases [17]. However, this method considers the detection in the gray scale, which cannot effectively explore the color information of the images and limits the accuracy. Moreover, the defects have different scales and representations. The statistic-based methods cannot effectively distinguish the diverse defects from each other and suffer from poor accuracy.

There are also works regarding the task in the spatial domain and utilizing some filter-based methods to detect and localize the defects. A Gabor filter was considered to explore the hole-like defects and achieved a good detection performance on different scales [18]. Hough transformation was also utilized to model the different kinds of defects and improve the robustness of the detection [19]. Edge information was also considered in the defect detection procedure [20]. Yang et al. utilized a convolution operation to explore the contour and edge information of the steel image and modeled the complex steel defects.

Recently, CNN has demonstrated its amazing performance on the objection detection task [21–23]. Most of the works are established based on the YOLO [9] series, fast RCNN [24] series or other object detection methodologies. Shi et al. developed an improved faster RCNN method for accurate steel surface defect detection [25]. Zhan proposed a bilaterally symmetric UNet for detection [26]. Yang and Guo also modified the YOLO network for detecting the defects [27]. Recently, there are also generative adversarial network (GAN)-based methods for defect detection [28–30].

## 2.2. Deep Learning for Classification and Object Detection

Deep learning has demonstrated great effectiveness in the object classification task, which is usually considered as the backbone of object detection methods. LeNet [31] was the first CNN-based method for image classification, which proved to be a success on hand-written digital number recognition. In 2012, AlexNet demonstrated its amazing performance on the ImageNet [32] competition with 62.5% accuracy, and won the first prize with a large improvement from the second prize method. After that, there are numerous classification networks with well-designed architectures and good classification performances. VGGNet [33] utilized a very deep network to improve the classification accuracy to 74.0%. ResNet [34] introduced the residual connection into image classification and achieved better performances than previous works with 78.4% accuracy. The residual connection has been widely developed in various network designs. DenseNet [35] provided a dense connection to build the network for better information transmission with 79.2% accuracy. Recently, there have been different modifications on ResNet and DenseNet to improve the classification performance. Zhang et al. proposed a multi-level residual network design and improved the network representation capacity [36]. Gao et al. developed a multi-scale backbone for ResNet, and proposed Res2Net for image classification [37]. Xie et al. brought the aggregated residual transformation from ResNet and named their new network ResNext [38].

Based on the classification backbones, there are numerous object detection networks achieving good performances. R-CNN [39] is the first CNN-based method for object detection, which utilized a CNN to explore the features and used SVM for classification. After that, fast R-CNN [40] modified the structure of R-CNN, and made it to be an end-to-end technology for better performance with fast speed. Faster R-CNN [41] further improved the fast R-CNN with higher speed and accuracy. Mask R-CNN [42] combined the object detection and the semantic segmentation.

The R-CNN series methods are two-stage detection technologies, which separate the localization and classification steps and become time-consuming. To boost the speed of detection, the YOLO series methods are proposed to meet the real-time requirement. YOLO-v1 [9] is the first YOLO-series method with restricted parameters and computational cost. To boost the precision, YOLO-v2 [10] used a new backbone and considered other modifications with faster speed. After that, YOLO-v3, v4 and v5 successively improved the performance with well-designed network components.

Except for R-CNN and YOLO, there are also other network architectures for object detection. SSD [43] added the multi-scale feature exploration to YOLO and achieved better performance. RetinaNet [44] used focal loss and regarded ResNet as the backbone to boost the accuracy. CenterNet [45] considered the object detection in an anchor-free style and jointly improved the speed and the accuracy. FCOS [46] also developed a fully convolutional network for one-stage object detection.

## 2.3. Deep Learning for Defect Detection

With the development of deep learning technology, numerous CNN-based methods have been especially designed for defect detections. The deep learning-based defect detection works usually concentrate on the effective network design and utilize well-established architectures to improve the network representation technology and boost the detection accuracy [47]. Wu et al. applied an SSD-based detection network for accurate

PCB defect detection [48]. An et al. developed an improved faster R-CNN network for fabric defect detection, which utilizes the VGG-16 backbone for feature extraction and builds a multi-scale feature pyramid model for the RPN network [49]. Luo et al. developed a decoupled two-stage network for the FPCB surface defect detection [50]. In their work, the localization and classification operations are decoupled into two different modules. Additionally, they established the multi-hierarchical aggregation block and the locally non-local block for boosting the network performance. With the elaborate network designs, their network achieves state-of-the-art performance with 91.45% mAP on FPCBs' defect detection. Guan et al. also developed an improved YOLOv5 network to detect the ceramic ring defects [51]. In their work, the attention mechanism was specially addressed into the YOLOv5 backbone and improved the detection performance. Their method achieves 89.9% mAP on the ceramic ring defect and achieves a state-of-the-art performance. Mo et al. proposed a weighted double-low-rank decomposition technology for fabric defect detection [52]. In contrast to other YOLO-based and R-CNN-based networks, their work regards the defect detection as an optimization-based problem and utilizes an alternating direction method of multipliers (ADMMs) to solve the task. With the new perspective and the novel methodology, their work achieves higher accuracy and better detection performance than other fabric defect detection methods. In contrast to the objective-oriented detection methods, Zeng et al. proposed a reference-based defect detection network for all defect detection tasks [53]. This method using a well-aligned template reference to estimate the potential defects of the input images.

Importantly, there are different deep learning-based works concentrating on the steel defect detection. YOLO and R-CNN are the most two popular baselines for developing the detection network. Su et al. developed an improved YOLO-v4 network for steel surface defect detection [54]. In their method, the channel attention mechanism was specially developed to capture the global information of the image feature. After that, an ICIoU loss function was introduced to replace the CIoU, which can more effectively solve the data imbalance issue. Their method achieves 78.63% mAP on the steel surface defect detection dataset and proves to be one of the state-of-the-art methods. Xie et al. also developed an improved faster R-CNN method for fast and accurate surface defect detection [55]. They modified the backbone of the faster R-CNN to better explore the image feature and achieve more accurate detection results. Beyond the YOLO and R-CNN series, there are also different technologies especially devised for steel surface defect detection. Tian et al. proposed a complementary adversarial network-driven surface defect detection for different types of the defects [56]. In their work, an encoding–decoding architecture was specially developed for image segmentation and the discriminator loss was considered for its better performance. Additionally, the dilated convolution and the edge detection are also considered in the network to effectively explore the image feature. Zhan also developed a bilaterally symmetric U-shaped network, dubbed BSU-Net, for effective surface defect detection [26]. In BSU-Net, an enhanced U-Net and a feature expanding network are combined to classify whether the image has defects. Cheng and Yu. considered the RetinaNet as the backbone, and embedded the channel attention mechanism and the adaptively spatial feature fusion into the detection procedure to boost the accuracy [57]. Guan et al. devised a U-shaped architecture to detect the defects, which used the VGG-19 as a feature extractor to extract the information. Furthermore, the structural similarity and the decision tree are utilized to evaluate the image quality [58]. Han et al. developed a two-stage edge reuse network embedding the saliency information into defect detection [59]. In their method, an edge-aware foreground–background integration module was especially devised to explore the saliency and further concentrate on the defect information.

#### 2.4. Attention Mechanism

Attention mechanism proves to be an effective component for CNN to boost the representation performance and improve the predicting results. In general, the attention mechanism can be separated into three different kinds: the channel-wise attention mecha-

nism [60], the spatial attention mechanism [61] and the non-local attention mechanism [62]. The channel-wise attention mechanism embeds the image features into a vector and gives different weights to different feature channels. In contrast to the channel-wise attention, the spatial attention mechanism finds the weights for every pixel of the image features. The non-local attention mechanism calculates the global relationship of the image feature, and utilizes the matrix multiplication operation to conduct the attention procedure. The attention mechanism has been widely used in different computer vision and image processing tasks, such as image super-resolution [60], image dehazing [63], object detection [61] and image segmentation [62].

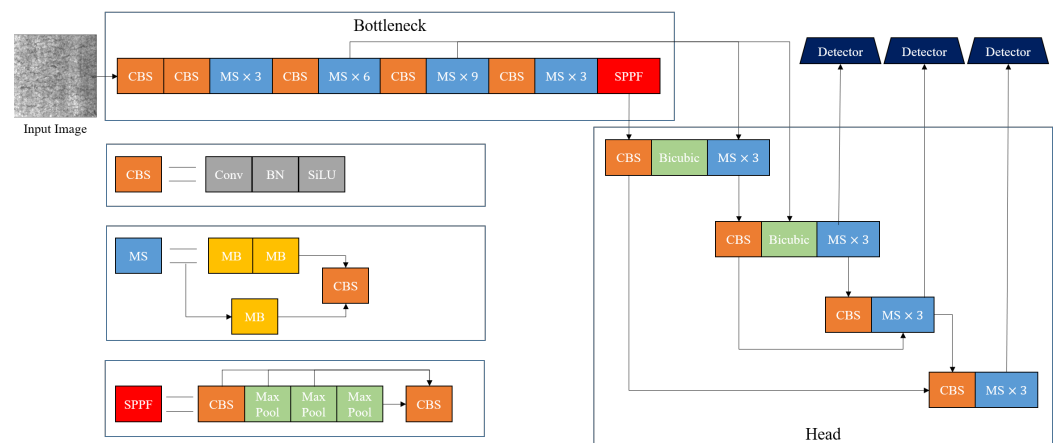
The attention mechanism is also widely considered in defect detection areas. Wang et al. used the spatial attention mechanism to detect the subway tunnel defects [64]. Li et al. devised a dynamic attention graph convolution mechanism for the point cloud defect detection [65]. Wu and Lu combined the spatial attention, channel attention and the non-local attention mechanisms for fabric defect detection and achieved a 91.6% mAP performance [66]. Chen et al. used the deformable convolution and the channel attention mechanism for building the strip steel surface defect detection network [67]. Peng et al. also developed a fabric defect detection network with both the spatial and channel attention mechanisms.

### 3. Method

In this section, we firstly introduce the design of the proposed network. Then, the multi-scale block and the spatial attention mechanism are described. Finally, we demonstrate the detailed implementation of the proposed network.

#### 3.1. Network Design

Figure 2 shows the network design of the proposed improved multi-scale YOLO-v5 method. This network is composed of three different components: the bottleneck, the head and the detector. The input image is firstly processed by the bottleneck to explore the multi-scale features. Then, the proposed features are aggregated and further processed by the head. Finally, the multi-scale features are sent to the detector for classification and localization.



**Figure 2.** Network design of the proposed improved multi-scale YOLO-v5 method.

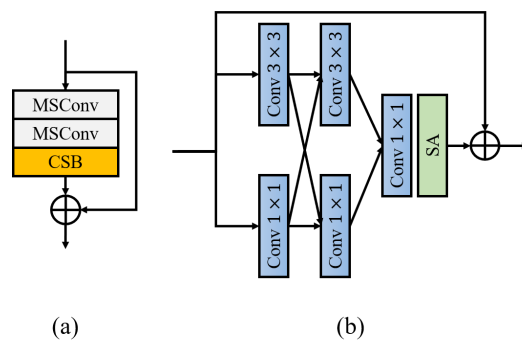
As shown in the figure, the bottleneck is composed of the combination of convolution, batch normalization and SiLU activation (CBS), the multi-scale sequence (MS) and the spatial pyramid pooling fusion (SPPF). There are five CBSs, fifteen MSs and one SPPF in the bottleneck. In MS, there are three multi-scale blocks (MBs) and one CBS for multi-scale feature fusion. The SPPF is composed of two CBSs and three max pooling (MaxPool) operations. The MaxPool operations explore the image feature in the spatial pyramid pooling fashion. Then, CBS combines and fuses the multi-scale features.

The head of the network is composed of four CBSs, twelve MSs and several bicubic operations to maintain the image resolution. The head combines features from the different stages of the bottleneck, and uses CBS and MSs for better multi-scale feature fusion. Finally, the multi-scale features of the head are sent to the detector for object detection and localization.

The detector follows the vanilla YOLO-v5 design [13], which regresses the bias of different anchors and localizes the objects. The detector contains three scales to effectively explore the small and large objects. For each scale, there are three anchors to localize the defects.

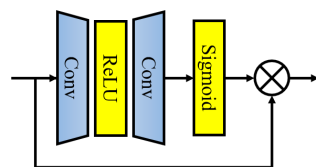
### 3.2. Design of the Multi-Scale Block and Spatial Attention

Figure 3a shows the design of the MB. There are two multi-scale convolutions (MSConv) to explore the hierarchical image information. After that, one CSB with skip connection builds the residual structure for better gradient transmission. Figure 3b demonstrates the design of MSConv. In the MSConv, two  $1 \times 1$  and two  $3 \times 3$  convolutions crossly process the image feature and explore the multi-scale information. After that, one  $1 \times 1$  convolution combines the features of two convolutions for information fusion and keeps the number of channels. A spatial attention (SA) mechanism is specifically developed to further concentrate on the defect information and improve the detection performance. Finally, a skip connection is introduced for better gradient transmission.



**Figure 3.** Design of the multi-scale block (MB): (a) Multi-scale block (MB); and (b) multi-scale convolution (MSConv).

Figure 4 shows the design of the SA. In the figure, we can find that the SA has two convolutions, one ReLU activation and one sigmoid activation. The convolutions decrease and increase the channel number symmetrically. The ReLU activation introduces the non-linearity to the attention exploration. Finally, the sigmoid activation introduces the non-negativity to the attention.



**Figure 4.** Design of the spatial attention (SA) mechanism.

The spatial attention mechanism follows an encoder–decoder design, which can effectively explore the spatial correlation of the input image feature. The sigmoid activation brings the non-negativity to the feature and gives higher weights to the detected areas, which helps boost the network representation capacity and improve the detection accuracy.

### 3.3. Implementation Details

Table 1 shows the parameter settings of the proposed improved multi-scale YOLO-v5 network. The component index follows the order in Figure 2. The scale of CBS means

decreasing the resolution of the image feature by  $s$  times, and the scale of bicubic means increasing the resolution by  $s$  times, where  $s$  is the scale.

**Table 1.** Parameter settings of the improved multi-scale YOLO-v5 network.

Component	Index	Operation	Number	Channel	Scale
Backbone	0	CBS	1	64	2
	1	CBS	1	128	2
	2	MS	3	128	1
	3	CBS	1	256	2
	4	MS	6	256	1
	5	CBS	1	512	2
	6	MS	9	512	1
	7	CBS	1	1024	2
	8	MS	3	1024	1
	9	SPPF	1	1024	1
Head	10	CBS	1	512	1
	11	Bicubic	1	-	2
	12	MS	3	512	1
	13	CBS	1	256	1
	14	Bicubic	1	-	2
	15	MS	3	256	1
	16	CBS	1	256	2
	17	MS	3	512	1
	18	CBS	1	512	2
	19	MS	3	1024	1

During the training phase, we used the same loss functions as the YOLO-v5, including the coordinate loss, the target confidence loss and the target classification loss. The weights and the implementation are entirely the same as the vanilla YOLO-v5 design for a fair comparison.

## 4. Experiment

### 4.1. Settings

We chose the NEU-DET [2] dataset to train and test our model. NEU-DET contains 1800 steel surface defect images with six typical defects: pitted surface, rolled-in scale, scratches, crazing, inclusion and patches. Among the images, we randomly chose 60% for training, 20% for validation and 20% for testing. We trained the network on one NVIDIA RTX 3080-Ti GPU. The batch size was chosen as 16. We updated the network for 100 epochs. The optimizer was chosen as Adam with a learning rate as  $10^{-3}$ .

The measurements of the performance are chosen as precision, recall and mean average precision (mAP). The precision and the recall are defined as

$$P = \frac{TP}{TP + FP'} \quad (1)$$

and

$$R = \frac{TP}{TP + FN'} \quad (2)$$

where  $TP$ ,  $FP$  and  $FN$  are the true positive, false positive and false negative samples, respectively.  $P$  is the precision and  $R$  is the recall.

### 4.2. Results

To demonstrate the effectiveness of our method, we mainly compared the improved version with two vanilla YOLO-v5 network settings: YOLOv5-s and YOLOv5-m. We firstly compared the computational complexity of different methods. Table 2 shows the parameters, GFLOPs and time costs of different methods. The GFLOPs are calculated by



processing one  $640 \times 640$  image. In the table, we can find that our method satisfies the real-time speed requirement and has the ability to process more than 190 images per second.

**Table 2.** Computational complexity comparisons among different methods.

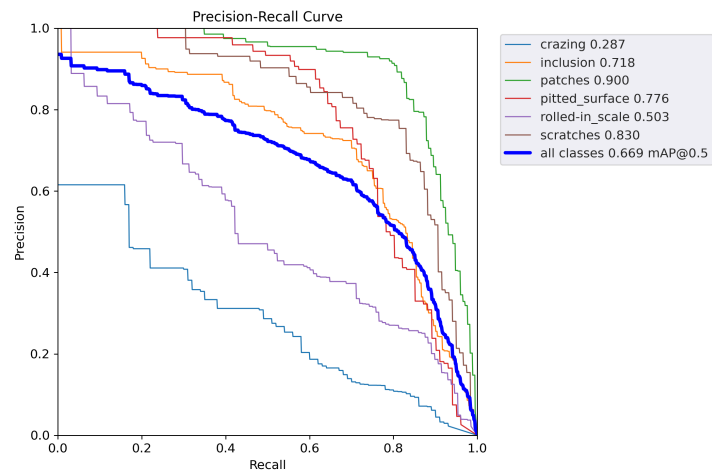
Method	Parameters (M)	GFLOPs	Time Cost (ms)	FPS
YOLO-v5s	7.02	15.8	1.8	555.56
YOLO-v5m	20.8	47.9	4.1	243.90
Ours	22.2	54.1	5.2	192.30

To demonstrate the effectiveness of our method, we compared it with YOLO-v7tiny [68], one of the state-of-the-art object detection methods. Table 3 shows the precision, recall, mAP50 and mAP50-95 comparisons among the different object detection methods. In the table, we can find that our network achieves the highest scores on all testing indicators. From this point of view, our method can effectively detect the defects of the steel surfaces. Figure 5 shows the PR-curve among different methods. In the figure, we can find that our method has a larger area under the curve (AUC), which denotes a better performance than the other methods. To further investigate the effectiveness of our method, we also demonstrated the precision, recall and F1 curves. Figure 6 shows the results of different indicators. We can find that our method has a good performance on different kinds of defects. Finally, we demonstrated the visualized results of the steel surface defect detection. Figure 7 shows the comparison between ground-truth and our prediction results. In the figure, we can find that our method can predict most of the defects on the steel and has a robust performance in terms of defects with different scales.

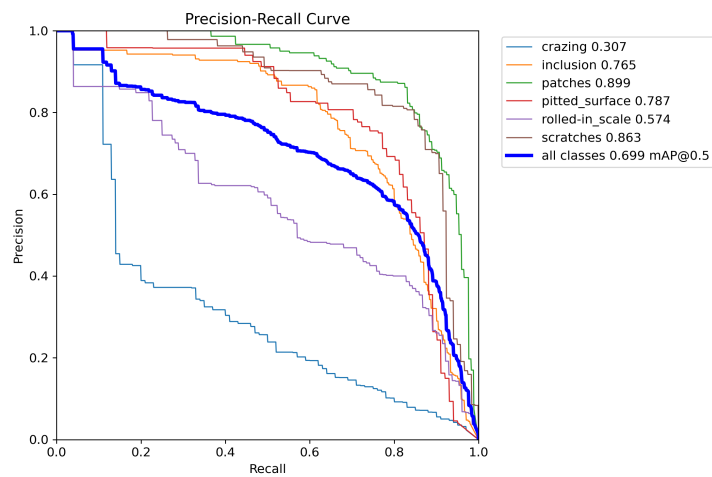
**Table 3.** Precision, recall, mAP50 and mAP50-95 comparisons among different methods.

Method	Indicator	Crazing	Inclusion	Patches	Pitted Surface	Rolled-in Scale	Scratches	All
YOLO-v5s	P	0.433	0.606	0.802	0.712	0.469	0.753	0.633
	R	0.010	0.758	0.849	0.723	0.680	0.814	0.597
	mAP50	0.287	0.718	0.900	0.776	0.574	0.830	0.669
	mAP50-95	0.089	0.330	0.554	0.406	0.257	0.415	0.334
YOLO-v5m	P	0.454	0.536	0.735	0.754	0.489	0.714	0.610
	R	0.140	0.833	0.884	0.759	0.430	0.873	0.695
	mAP50	0.307	0.765	0.899	0.787	0.503	0.863	0.699
	mAP50-95	0.099	0.384	0.571	0.451	0.212	0.444	0.368
YOLO-v7tiny	P	1.000	0.538	0.751	0.628	0.384	0.624	0.654
	R	0.000	0.738	0.824	0.574	0.342	0.746	0.537
	mAP50	0.168	0.659	0.835	0.626	0.332	0.713	0.555
	mAP50-95	0.036	0.282	0.451	0.270	0.101	0.300	0.240
Ours	P	0.573	0.595	0.759	0.743	0.505	0.766	0.657
	R	0.180	0.819	0.890	0.772	0.703	0.864	0.705
	mAP50	0.345	0.768	0.898	0.825	0.616	0.868	0.720
	mAP50-95	0.114	0.373	0.576	0.451	0.277	0.440	0.372

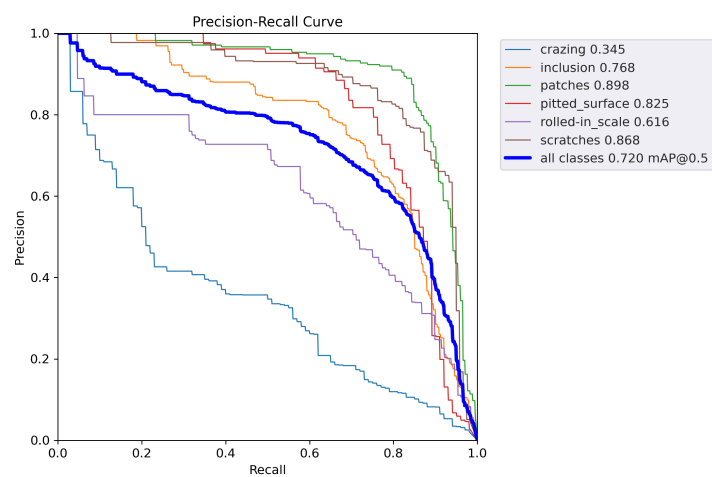
The performance gain comes from the well-designed network architecture. In Table 2, our method has similar parameters, GFLOPs and time costs to YOLO-v5m. In contrast, the performance of our method is superior to YOLO-v5m. It should be noticed that YOLO-v5m is a larger version of YOLO-v5s, whose performance improvement is limited. From this point of view, the performance gain comes from the new architecture rather than the larger network.



(a) YOLO-v5s

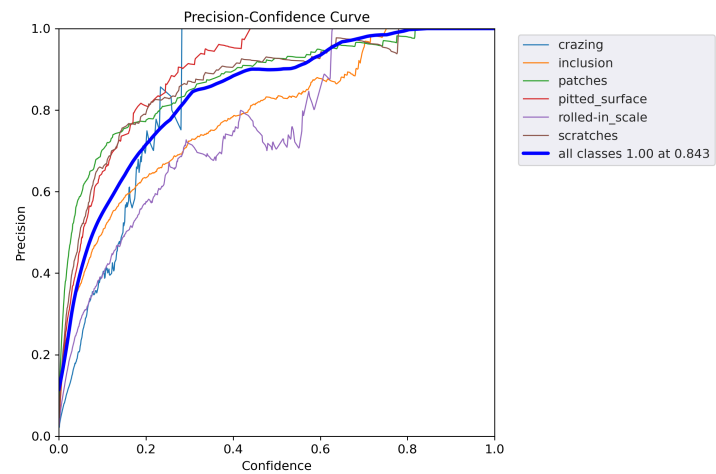


(b) YOLO-v5m

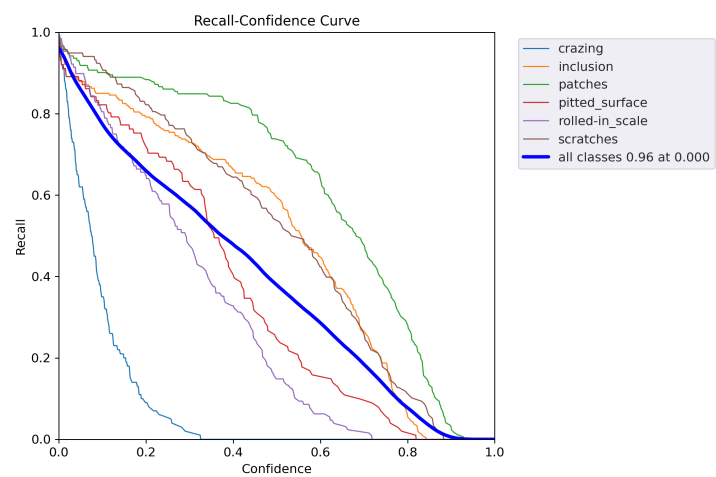


(c) Ours

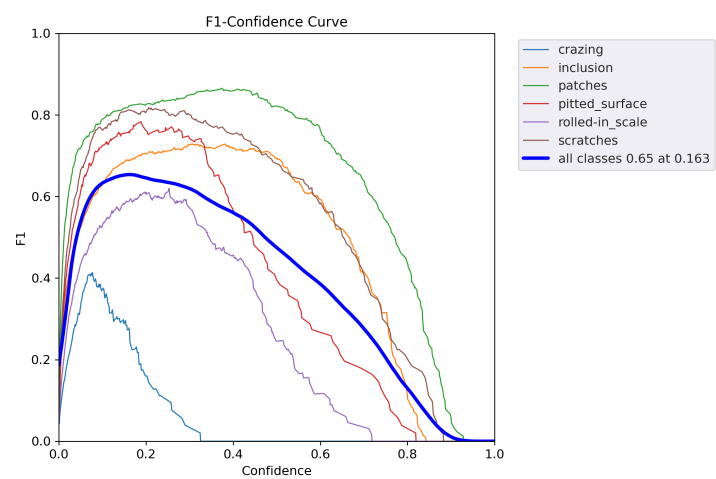
Figure 5. PR curve comparisons among different methods.



(a) Precision

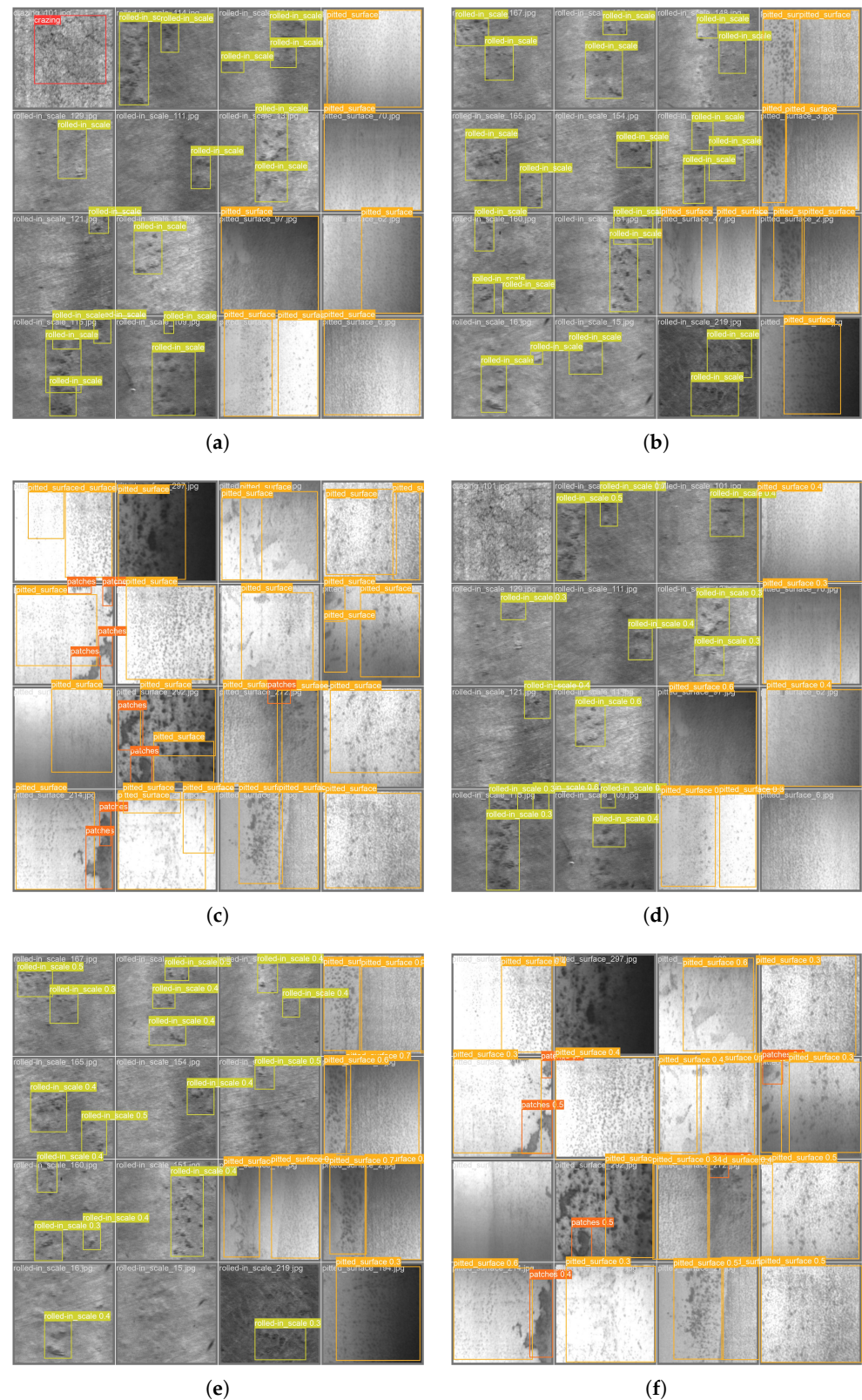


(b) Recall



(c) F1

**Figure 6.** Precision, recall and F1 curves of our method.



**Figure 7.** Visualized results of the steel surface defect detection: (a–c) Groundtruth. (d–f) Prediction results. Zoom-up for better view.

It should be noticed that the best mAP50 performance of Table 3 is of approximately 0.72, which is lower than other reports. This is because we used an entirely different data organization protocol from other papers. In our work, the NEU-DET dataset is split by 60%,

20% and 20% for training, validation and testing, respectively. The amount of training data is much smaller than in other works for ensuring the generation performance. To fairly compare the effectiveness of different methods, we re-trained different methods under the same protocol, the results of which are reliable for measuring the performances.

## 5. Conclusions

In this paper, we proposed an improved multi-scale YOLO-v5 network for steel surface defect detection. To focus on diverse defects at different scales, we developed a multi-scale block to effectively explore the defects with different resolutions. To further improve the network performance and concentrate more on the defect areas, we developed a spatial attention mechanism to give higher weights to abnormal information. The experimental results show that the improved multi-scale YOLO-v5 network can effectively detect different kinds and scales of defects and satisfies the real-time speed requirement.

**Author Contributions:** Conceptualization, X.L. and L.W.; methodology, L.W.; software, X.L.; validation, J.M., W.S. and H.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work is supported in part by the Cooperation Innovation Plan of Yingkou for Enterprise and Doctor (QB-2019-10, 2022-13), the Liaoning Science and Technology Joint Fund (2020-YKLH-26, 2021-YKLH-19), the Foundation of Liaoning Key Laboratory of Chemical Additive Synthesis and Separation (ZJNK2109), the Program for Excellent Talents of Science and Technology in Yingkou Institute of Technology (RC201902) and the Liaoning Province's Science and Technology Plan (Major Project of "Jiebangguashuai" (2022JH1/10400009).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The dataset is public in the reference [2].

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Luo, Q.; Fang, X.; Liu, L.; Yang, C.; Sun, Y. Automated Visual Defect Detection for Flat Steel Surface: A Survey. *IEEE Trans. Instrum. Meas.* **2020**, *69*, 626–644. [\[CrossRef\]](#)
2. Lv, X.; Duan, F.; Jiang, J.J.; Fu, X.; Gan, L. Deep Metallic Surface Defect Detection: The New Benchmark and Detection Network. *Sensors* **2020**, *20*, 1562. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Amin, D.; Akhter, S. Deep Learning-Based Defect Detection System in Steel Sheet Surfaces. In Proceedings of the 2020 IEEE Region 10 Symposium (TENSymp), Dhaka, Bangladesh, 5–7 June 2020; pp. 444–448. [\[CrossRef\]](#)
4. Caleb, P.; Steuer, M. Classification of surface defects on hot rolled steel using adaptive learning methods. In Proceedings of the KES'2000, Fourth International Conference on Knowledge-Based Intelligent Engineering Systems and Allied Technologies, Proceedings (Cat. No. 00TH8516), Brighton, UK, 30 August–1 September 2000; Volume 1, pp. 103–108. [\[CrossRef\]](#)
5. Choi, K.; Koo, K.; Lee, J.S. Development of Defect Classification Algorithm for POSCO Rolling Strip Surface Inspection System. In Proceedings of the 2006 SICE-ICASE International Joint Conference, Busan, Republic of Korea, 18–21 October 2006; pp. 2499–2502. [\[CrossRef\]](#)
6. Rautkorpi, R.; Iivarinen, J. Content-based image retrieval of Web surface defects with PicSOM. In Proceedings of the 2004 IEEE International Joint Conference on Neural Networks (IEEE Cat. No. 04CH37541), Budapest, Hungary, 25–29 July 2004; Volume 3, pp. 1863–1867. [\[CrossRef\]](#)
7. Tunali, M.M.; Yildiz, A.; Çakar, T. Steel Surface Defect Classification Via Deep Learning. In Proceedings of the 2022 7th International Conference on Computer Science and Engineering (UBMK), Diyarbakir, Turkey, 14–16 September 2022; pp. 485–489. [\[CrossRef\]](#)
8. He, Y.; Song, K.; Meng, Q.; Yan, Y. An End-to-End Steel Surface Defect Detection Approach via Fusing Multiple Hierarchical Features. *IEEE Trans. Instrum. Meas.* **2020**, *69*, 1493–1504. [\[CrossRef\]](#)
9. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788. [\[CrossRef\]](#)
10. Redmon, J.; Farhadi, A. YOLO9000: Better, Faster, Stronger. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 6517–6525. [\[CrossRef\]](#)
11. Chen, D.; Ju, Y. SAR ship detection based on improved YOLOv3. In Proceedings of the IET International Radar Conference (IET IRC 2020), Virtual, 4–6 November 2020; Volume 2020, pp. 929–934. [\[CrossRef\]](#)

12. Liu, T.; Chen, S. YOLOv4-DCN-based fabric defect detection algorithm. In Proceedings of the 2022 37th Youth Academic Annual Conference of Chinese Association of Automation (YAC), Beijing, China, 27–29 May 2022; pp. 710–715. [[CrossRef](#)]
13. Li, Y.; Cheng, R.; Zhang, C.; Chen, M.; Ma, J.; Shi, X. Sign language letters recognition model based on improved YOLOv5. In Proceedings of the 2022 9th International Conference on Digital Home (ICDH), Guangzhou, China, 28–30 October 2022; pp. 188–193. [[CrossRef](#)]
14. He, B.; Zhuo, J.; Zhuo, X.; Peng, S.; Li, T.; Wang, H. Defect detection of printed circuit board based on improved YOLOv5. In Proceedings of the 2022 International Conference on Artificial Intelligence and Computer Information Technology (AICIT), Yichang, China, 16–18 September 2022; pp. 1–4. [[CrossRef](#)]
15. Zheng, L.; Wang, X.; Wang, Q.; Wang, S.; Liu, X. A Fabric Defect Detection Method Based on Improved YOLOv5. In Proceedings of the 2021 7th International Conference on Computer and Communications (ICCC), Chengdu, China, 10–13 December 2021; pp. 620–624. [[CrossRef](#)]
16. Djukic, D.; Spuzic, S. Statistical discriminator of surface defects on hot rolled steel. In Proceedings of the Image and Vision Computing New Zealand 2007, Hamilton, New Zealand, 5–7 December 2007; pp. 158–163.
17. Wang, Y.; Xia, H.; Yuan, X.; Li, L.; Sun, B. Distributed defect recognition on steel surfaces using an improved random forest algorithm with optimal multi-feature-set fusion. *Multimed. Tools Appl.* **2018**, *77*, 16741–16770. [[CrossRef](#)]
18. Choi, D.c.; Jeon, Y.J.; Kim, S.H.; Moon, S.; Yun, J.P.; Kim, S.W. Detection of pinholes in steel slabs using Gabor filter combination and morphological features. *ISIJ Int.* **2017**, *57*, 1045–1053. [[CrossRef](#)]
19. Sharifzadeh, M.; Amirfattahi, R.; Sadri, S.; Alirezaee, S.; Ahmadi, M. Detection of steel defect using the image processing algorithms. In Proceedings of the International Conference on Electrical Engineering, Military Technical College, Dhaka, Bangladesh, 20–22 December 2008; Volume 6, pp. 1–7.
20. Yang, J.; Li, X.; Xu, J.; Cao, Y.; Zhang, Y.; Wang, L.; Jiang, S. Development of an optical defect inspection algorithm based on an active contour model for large steel roller surfaces. *Appl. Opt.* **2018**, *57*, 2490–2498. [[CrossRef](#)] [[PubMed](#)]
21. Liu, Y.; Wang, J.; Yu, H.; Li, F.; Yu, L.; Zhang, C. Surface Defect Detection of Steel Products Based on Improved YOLOv5. In Proceedings of the 2022 41st Chinese Control Conference (CCC), Hefei, China, 25–27 July 2022; pp. 5794–5799.
22. Cheng, Y.; Wang, S. Improvements to YOLOv4 for Steel Surface Defect Detection. In Proceedings of the 2022 5th International Conference on Intelligent Autonomous Systems (ICoIAS), Dalian, China, 23–25 September 2022; pp. 48–53.
23. Zhang, Y.; Xiao, F.; Tian, P. Surface defect detection of hot rolled steel strip based on image compression. In Proceedings of the 2020 International Conference on Computer Network, Electronic and Automation (ICCNEA), Xi'an, China, 25–27 September 2020; pp. 149–153.
24. Ullah, A.; Xie, H.; Farooq, M.O.; Sun, Z. Pedestrian detection in infrared images using fast RCNN. In Proceedings of the 2018 Eighth International Conference on Image Processing Theory, Tools and Applications (IPTA), Xi'an, China, 7–10 November 2018; pp. 1–6.
25. Shi, X.; Zhou, S.; Tai, Y.; Wang, J.; Wu, S.; Liu, J.; Xu, K.; Peng, T.; Zhang, Z. An Improved Faster R-CNN for Steel Surface Defect Detection. In Proceedings of the 2022 IEEE 24th International Workshop on Multimedia Signal Processing (MMSP), Shanghai, China, 26–28 September 2022; pp. 1–5.
26. Xinzi, Z. BSU-net: A surface defect detection method based on bilaterally symmetric U-Shaped network. In Proceedings of the 2020 5th International Conference on Mechanical, Control and Computer Engineering (ICMCCE), Harbin, China, 25–27 December 2020; pp. 1771–1775.
27. Yang, N.; Guo, W. Application of Improved YOLOv5 Model for Strip Surface Defect Detection. In Proceedings of the 2022 Global Reliability and Prognostics and Health Management (PHM-Yantai), Yantai, China, 13–16 October 2022; pp. 1–5.
28. Xu, L.; Tian, G.; Zhang, L.; Zheng, X. Research of surface defect detection method of hot rolled strip steel based on generative adversarial network. In Proceedings of the 2019 Chinese Automation Congress (CAC), Hangzhou, China, 22–24 November 2019; pp. 401–404.
29. Liu, K.; Li, A.; Wen, X.; Chen, H.; Yang, P. Steel surface defect detection using GAN and one-class classifier. In Proceedings of the 2019 25th International Conference on Automation and Computing (ICAC), Lancaster, UK, 5–7 September 2019; pp. 1–6.
30. Wen, L.; Wang, Y.; Li, X. A new Cycle-consistent adversarial networks with attention mechanism for surface defect classification with small samples. *IEEE Trans. Ind. Inform.* **2022**, *18*, 8988–8998. [[CrossRef](#)]
31. Al-Jawfi, R. Handwriting Arabic character recognition LeNet using neural network. *Int. Arab. J. Inf. Technol.* **2009**, *6*, 304–309.
32. Deng, J.; Dong, W.; Socher, R.; Li, L.J.; Li, K.; Li, F.-F. Imagenet: A large-scale hierarchical image database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 248–255.
33. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
34. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
35. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708.
36. Zhang, K.; Sun, M.; Han, T.X.; Yuan, X.; Guo, L.; Liu, T. Residual networks of residual networks: Multilevel residual networks. *IEEE Trans. Circuits Syst. Video Technol.* **2017**, *28*, 1303–1314. [[CrossRef](#)]
37. Gao, S.H.; Cheng, M.M.; Zhao, K.; Zhang, X.Y.; Yang, M.H.; Torr, P. Res2net: A new multi-scale backbone architecture. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *43*, 652–662. [[CrossRef](#)] [[PubMed](#)]

38. Xie, S.; Girshick, R.; Dollár, P.; Tu, Z.; He, K. Aggregated residual transformations for deep neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1492–1500.
39. Bharati, P.; Pramanik, A. Deep learning techniques—R-CNN to mask R-CNN: A survey. In *Computational Intelligence in Pattern Recognition: Proceedings of CIPR 2019*; Springer, Singapore, 2020; pp. 657–668.
40. Girshick, R. Fast r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1440–1448.
41. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. In Proceedings of the Advances in neural Information Processing Systems, Montreal, QC, Canada, 7–12 December 2015; Volume 28.
42. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2961–2969.
43. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. Ssd: Single shot multibox detector. In *Lecture Notes in Computer Science, Proceedings of the Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016, Proceedings, Part I 14*; Springer: Berlin/Heidelberg, Germany, 2016; pp. 21–37.
44. Lin, T.Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal loss for dense object detection. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2980–2988.
45. Zhou, X.; Wang, D.; Krähenbühl, P. Objects as points. *arXiv* **2019**, arXiv:1904.07850.
46. Tian, Z.; Shen, C.; Chen, H.; He, T. Fcos: Fully convolutional one-stage object detection. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 9627–9636.
47. Liu, G. Surface Defect Detection Methods Based on Deep Learning: A Brief Review. In Proceedings of the 2020 2nd International Conference on Information Technology and Computer Application (ITCA), Guangzhou, China, 17–19 December 2020; pp. 200–203. [[CrossRef](#)]
48. Wu, X.; Ge, Y.; Zhang, Q.; Zhang, D. PCB Defect Detection Using Deep Learning Methods. In Proceedings of the 2021 IEEE 24th International Conference on Computer Supported Cooperative Work in Design (CSCWD), Dalian, China, 5–7 May 2021; pp. 873–876. [[CrossRef](#)]
49. An, M.; Wang, S.; Zheng, L.; Liu, X. Fabric defect detection using deep learning: An Improved Faster R-approach. In Proceedings of the 2020 International Conference on Computer Vision, Image and Deep Learning (CVIDL), Nanchang, China, 15–17 May 2020; pp. 319–324. [[CrossRef](#)]
50. Luo, J.; Yang, Z.; Li, S.; Wu, Y. FPCB Surface Defect Detection: A Decoupled Two-Stage Object Detection Framework. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 5012311. [[CrossRef](#)]
51. Guan, S.; Wang, X.; Wang, J.; Yu, Z.; Wang, X.; Zhang, C.; Liu, T.; Liu, D.; Wang, J.; Zhang, L. Ceramic ring defect detection based on improved YOLOv5. In Proceedings of the 2022 3rd International Conference on Computer Vision, Image and Deep Learning and International Conference on Computer Engineering and Applications (CVIDL and ICCEA), Changchun, China, 20–22 May 2022; pp. 115–118. [[CrossRef](#)]
52. Mo, D.; Wong, W.K.; Lai, Z.; Zhou, J. Weighted Double-Low-Rank Decomposition With Application to Fabric Defect Detection. *IEEE Trans. Autom. Sci. Eng.* **2021**, *18*, 1170–1190. [[CrossRef](#)]
53. Zeng, Z.; Liu, B.; Fu, J.; Chao, H. Reference-Based Defect Detection Network. *IEEE Trans. Image Process.* **2021**, *30*, 6637–6647. [[CrossRef](#)] [[PubMed](#)]
54. Su, Y.; Zhang, Q.; Deng, Y.; Luo, Y.; Wang, X.; Zhong, P. Steel Surface Defect Detection Algorithm based on Improved YOLOv4. In Proceedings of the 2022 IEEE 5th Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC), Chongqing, China, 16–18 December 2022; Volume 5, pp. 1425–1429. [[CrossRef](#)]
55. Xie, Q.; Zhou, W.; Tan, H.; Wang, X. Surface Defect Recognition in Steel Plates Based on Improved Faster R-CNN. In Proceedings of the 2022 41st Chinese Control Conference (CCC), Hefei, China, 25–27 July 2022; pp. 6759–6764. [[CrossRef](#)]
56. Tian, S.; Huang, P.; Ma, H.; Wang, J.; Zhou, X.; Zhang, S.; Zhou, J.; Huang, R.; Li, Y. CASDD: Automatic Surface Defect Detection Using a Complementary Adversarial Network. *IEEE Sens. J.* **2022**, *22*, 19583–19595. [[CrossRef](#)]
57. Cheng, X.; Yu, J. RetinaNet With Difference Channel Attention and Adaptively Spatial Feature Fusion for Steel Surface Defect Detection. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 2503911. [[CrossRef](#)]
58. Guan, S.; Lei, M.; Lu, H. A Steel Surface Defect Recognition Algorithm Based on Improved Deep Learning Network Model Using Feature Visualization and Quality Evaluation. *IEEE Access* **2020**, *8*, 49885–49895. [[CrossRef](#)]
59. Han, C.; Li, G.; Liu, Z. Two-Stage Edge Reuse Network for Salient Object Detection of Strip Steel Surface Defects. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 5019812. [[CrossRef](#)]
60. Fan, Z.; Dan, T.; Yu, H.; Liu, B.; Cai, H. Single Fundus Image Super-Resolution Via Cascaded Channel-Wise Attention Network. In Proceedings of the 2020 42nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Montreal, QC, Canada, 20–24 July 2020; pp. 1984–1987. [[CrossRef](#)]
61. Lu, H.; Chen, X.; Zhang, G.; Zhou, Q.; Ma, Y.; Zhao, Y. Scanet: Spatial-channel Attention Network for 3D Object Detection. In Proceedings of the ICASSP 2019—2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 12–17 May 2019; pp. 1992–1996. [[CrossRef](#)]
62. Guo, L.; Chen, L.; Philip Chen, C.L.; Li, T.; Zhou, J. Clustering based Image Segmentation via Weighted Fusion of Non-local and Local Information. In Proceedings of the 2018 International Conference on Security, Pattern Analysis, and Cybernetics (SPAC), Jinan, China, 14–17 December 2018; pp. 299–303. [[CrossRef](#)]

63. Zhou, J.; Leong, C.T.; Li, C. Multi-Scale and Attention Residual Network for Single Image Dehazing. In Proceedings of the 2021 6th International Conference on Intelligent Computing and Signal Processing (ICSP), Xi'an, China, 9–11 April 2021; pp. 483–487. [[CrossRef](#)]
64. Wang, A.; Togo, R.; Ogawa, T.; Haseyama, M. Multi-scale Defect Detection from Subway Tunnel Images with Spatial Attention Mechanism. In Proceedings of the 2022 IEEE International Conference on Consumer Electronics—Taiwan, Taipei, Taiwan, China, 6–8 July 2022; pp. 305–306. [[CrossRef](#)]
65. Li, Y.; Zhang, R.; Li, H.; Shao, X. Dynamic Attention Graph Convolution Neural Network of Point Cloud Segmentation for Defect Detection. In Proceedings of the 2020 IEEE International Conference on Artificial Intelligence and Information Systems (ICAIS), Dalian, China, 20–22 March 2020; pp. 18–23. [[CrossRef](#)]
66. Wu, X.; Lu, D. Parallel attention network based fabric defect detection. In Proceedings of the 2022 IEEE 5th Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC), Chongqing, China, 16–18 December 2022; Volume 5, pp. 1015–1020. [[CrossRef](#)]
67. Chen, H.; Du, Y.; Fu, Y.; Zhu, J.; Zeng, H. DCAM-Net: A Rapid Detection Network for Strip Steel Surface Defects Based on Deformable Convolution and Attention Mechanism. *IEEE Trans. Instrum. Meas.* **2023**, *72*, 5005312. [[CrossRef](#)]
68. Hong, X.; Wang, F.; Ma, J. Improved YOLOv7 Model for Insulator Surface Defect Detection. In Proceedings of the 2022 IEEE 5th Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC), Chongqing, China, 16–18 December 2022; Volume 5, pp. 1667–1672. [[CrossRef](#)]

**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.