



Article ID 1007-1202(2022)06-0489-10

DOI <https://doi.org/10.1051/wujns/2022276489>

Automatic Detection of Weld Defects in Pressure Vessel X-Ray Image Based on CNN

□ XIAO Wenkai^{1,2}, FENG Xiang^{1,2},
NAN Shuiyu³, ZHANG Linlin³

1. Department of Computer Science and Engineering, East China University of Science and Technology, Shanghai 200237, China;

2. Shanghai Engineering Research Center of Smart Energy, Shanghai 201103, China;

3. Shanghai Aino Industrial Technology CO., LTD, Shanghai 201612, China

© Wuhan University 2022

Abstract: The visual automatic detection method based on artificial intelligence has attracted more and more attention. In order to improve the performance of weld nondestructive defect detection, we propose DRepDet (Dilated RepPoints Detector). First, we analyze the weld defect dataset in detail and summarize the distribution characteristics of weld defect data, that is, the defect scale is very different and the aspect ratio distribution range is large. Second, according to the distribution characteristics of defect data, we design DResBlock module, and introduce dilated convolution with different dilated rates in the process of feature extraction to expand the receptive field and improve the detection performance of large-scale defects. Based on DResBlock and anchor-free detection framework RepPoints, we design DRepDet. Extensive experiments show that our proposed detector can detect 7 types of defects. When using combined dilated rate convolution network in detection, the AP50 and Recall50 of big defects are improved by 3.1% and 3.3% respectively, while the performance of small defects is not affected, almost the same or slightly improved. The final performance of the whole network is improved a large margin, with 6% AP50 and 4.2% Recall50 compared with Cascade R-CNN and 1.4% AP50 and 2.9% Recall50 compared with RepPoints.

Key words: nondestructive testing; depth learning; weld defect detection; convolutional neural networks; dilated convolution

CLC number: TP 183

Received date: 2022-09-28

Biography: XIAO Wenkai, male, Ph. D. candidate, Senior Engineer, research direction: artificial intelligence. E-mail: xiaowk@shenergy.com.cn

0 Introduction

In the manufacturing process of pressure vessels, many components need to be welded. In the process of welding, due to physical environment or human error, the welded joint will form defects at the weld, resulting in potential safety hazards. In order to ensure the quality and safety of welded parts and prevent accidents, defect detection of welded parts becomes very important. At present, non destructive testing (NDT) methods commonly used in industry include X-ray testing, ultrasonic testing, magnetic particle testing, eddy current testing, etc. According to different image characteristics, welding defects are generally divided into cracks, lack of penetration, lack of fusion, porosity and so on. Due to the different detection standards, the classification of defects will be different. For example, the strip and circle defects in the Chinese National Standard (NBT 47013.2-2015)^[1] include porosity and slag inclusions in the European Standard (EN ISO5817)^[2], but the definitions are different. In the current actual production, the main audit method is to manually analyze the weld image, judge whether there are defects and the type, location and size of defects based on experience, so as to evaluate the welding quality and give the corresponding rating. The manual evaluation method is affected by the testing staff's level, experience, fatigue and other human factors and external conditions, which is inefficient, unreliable and inconsistent.

In recent years, with the development of internet and the continuous improving of Graphics Processing Unit (GPU) hardware, computer vision based on deep learning has developed rapidly and gradually been the

dominant method in the field of image processing and pattern recognition. The defect detection method of industrial automation based on computer vision has also shown an explosive growth. Classification^[3-6], target detection^[7-10], segmentation^[11,12] and other methods in the field of computer vision can be used for industrial defect detection. Ref. [13] used a classification network based on convolution neural network to detect cloth defects; Ref. [14] employed the target detection algorithm to detect whether the fasteners of overhead catenary system support are missing; Ref. [15] used a semantic segmentation algorithm to detect metal surface defects. Considering that the annotating of segmentation is expensive and it is difficult to obtain enough annotation data, our work uses the object detection algorithm to detect weld defects.

Most of the object detection algorithms take the Microsoft Common Objects in Context (MS COCO)^[16] benchmark as the target dataset and optimize the algorithms on COCO benchmark. However, the X-ray weld images of pressure vessel are quite different from the COCO benchmark in terms of image resolution and defect scale. Therefore, the performance of many vanilla object detection algorithms on weld defect detection dataset has decreased. In order to solve the problems of low efficiency, unreliability and poor consistency of manual evaluation, we have done a lot of research on automatic defect detection, and statistically analyzed the weld defect dataset. On the basis of these research and analysis, we propose an automatic defect detection algorithm based on Convolutional Neural Network (CNN), which is mainly aimed at X-ray image detection. The reference standard is European Standard (EN ISO5817). The proposed detector can identify defects include porosity, slag inclusion, cracks, lack of penetration, lack of fusion, intensive dispersed and intensive chain, while most other automatic weld defect detectors can only detect no more than five kinds of defect.

The main contributions of our work are as follows:

- 1) We analyze the characteristics of the weld image and the weld defects thoroughly.
- 2) According to the characteristics of the weld defects to be detected, we propose DRepDet(dilation Rep-Points Detector) which can locate and identify a variety of defects including porosity, slag inclusions, cracks, lack of penetration, lack of fusion, defect intensive dispersed and intensive chain.
- 3) Extensive experiments are carried out on our pri-

vate weld image dataset to validate the effectiveness of the proposed algorithm. The results show that the proposed DRepDet can effectively improve the precision and recall of large defects while the precision and recall of small defects are not affected.

1 Related Work

1.1 Object Detection

The object detection task in computer vision needs to locate the object to be detected in a picture, and give the size and classification of the object. The anchor-based algorithm needs to preset the anchor box size and aspect ratio according to the statistical information of the ground truth bounding box. When the scale range of the object changes greatly, it can not adapt to the object well. RetinaNet^[17], Faster R-CNN^[9] and Cascade R-CNN^[10] are all such algorithms; The anchor-free algorithm directly regresses the center point, corner point or bounding point used to calculate the bounding box from the feature points, and then calculates the bounding box by clustering or transformation algorithm. CornerNet^[18] first predicted the corner points in the upper left corner and the lower right corner, and then used a special clustering method to obtain the bounding box of the target. ExtremeNet^[19] used segmentation supervision to locate the extreme point of the target in the *X-Y* direction. Rep-Points^[20] predicted a set of point representations from each feature map point, generated a pseudo bounding box from this set of point representations, and then used the ground truth box for semi-supervised learning. Whether the algorithm is based on anchor or the algorithm without anchor, the scale distribution of detection object/defect is very important for the selection and design of detection algorithm. Therefore, we conduct a multi-dimensional detailed analysis of weld defect data set.

1.2 Receptive Field

Receptive field is a basic concept of deep convolution neural network. It is a region of the input on which the convolution network depends. The value outside this region will not affect the output value. Ref. [21] proposed that the effective receptive field is smaller than the theoretical receptive field and follows the Gaussian distribution. The receptive field must be large enough to cover the whole relevant image area. Considering both efficiency and accuracy, dilated convolution is often

used to expand the receptive field in object detection and segmentation tasks. Deeplab series^[22,23] respectively explored the dilated convolution of different structures (including width, depth, etc.) to improve the accuracy of semantic segmentation. Based on the Inception^[24] structure, RFBNet^[25] used different scales of conventional convolution and dilated convolution on each branch to generate different receptive fields through different convolution kernel of conventional convolution. Ref. [26] used Kronecker revolution to solve the problem of local information loss in dilated convolution. Inspired by these works and combined with the distribution characteristics of weld defects, we design the Dilated RepPoints Detector (DRepDet). DRepDet uses dilated convolution to expand the receptive field, and uses different dilated rates according to different aspect ratios of defects to better solve the problem of huge difference in aspect ratio.

1.3 Defect Detection

With the rapid development of deep learning and computer vision technology, many advanced algorithms have been proposed for industrial defect detection. Ref. [27] proposed an automatic defect detection framework for aluminum conductor composite core (ACCC) based on image classification network, which takes inception ResNet as the backbone network. Based on the mask region-based CNN architecture^[28], an X-ray image casting defect detection system was constructed. In order to improve the accuracy of porosity defect detection, Ref. [29] proposed a semantic segmentation network based on encoder-decoder structure to recognize porosity defects at the pixel level. Ref. [30] used a three-stage system based on AdaBoost (defect extraction, defect detec-

tion and defect identification) to identify five weld defects (crack, lack of penetration, lack of fusion, circle and strip). The real defect detection scene of industrial site needs to detect more than 10 kinds of defects. In order to make the automatic defect system more suitable for the field environment, we design and implement a detection system that could detect 7 kinds of defects, including porosity, slag inclusion, cracks, lack of penetration, lack of fusion, intensive dispersed and intensive chain. As far as we know, this is a Liquefied Natural Gas (LNG) weld defect automatic detector with the most types of defects at present.

2 Problem Description

Compared with the images in COCO dataset, the situation of weld images is more complicated, mainly with differences in the following aspects: 1) In order to detect defects in more complex weld images, most of the images are 300 dpi-570 dpi, which leads to higher resolution images with size about 3 000×2 000-8 000×2 000; 2) The scale of weld defects varies greatly, the smallest defect is less than 10×10, and the largest defect is about 7 000×300; 3) The height/width ratio of weld defects ranges from 1:1 to 91:1, most of which are 1:1 and 2:1, as shown in Fig.1; 4) The edge line between the defect and the background is vague, as shown in Fig.2 (d), which makes the identification of defects more difficult. This paper mainly considers the scale difference and aspect ratio of weld defects, and designs a convolution network DRepDet suitable for weld defects to solve the problem of difficult identification of large-scale and abnormal aspect ratio defects. In Section 4.2, we conduct a

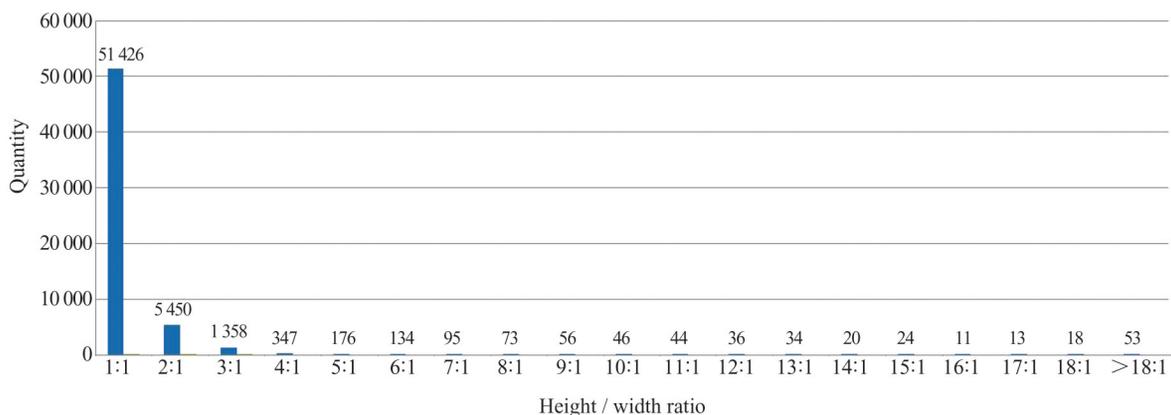


Fig. 1 Statistics of height/width ratio of weld defects

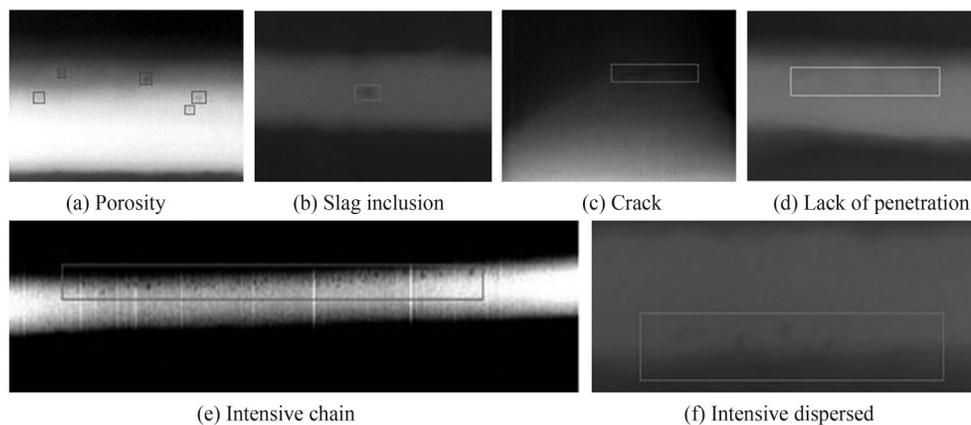


Fig. 2 Examples of weld defects

large number of experiments to verify the effectiveness of our proposed DRepDet.

There are several ways to expand the receptive field: 1) Make the network deeper by stacking more layers, which will increase the amount of parameters and calculation; 2) Make more down sampling, which will reduce the resolution of the feature map, and low-resolution features will reduce the accuracy of object positioning and recognition; 3) Use dilated convolution, which can effectively avoid the shortcomings of the first two methods, but the grid problem will occur. Table 1 lists the comparison of down sampling rate, receptive field, parameter and calculation complexity of the two models with different depths of the classical neural network ResNet in computer vision. Table 1 shows that increasing the depth of the model will greatly increase the number of parameters and calculations of the model, which has high requirements for memory and computing power, and is not applicable in industrial scenarios. Too much down sampling will lead to the loss of a large amount of detail information, which is not friendly for the detection of small defects, and there are a large number of small porosity defects in the weld defects. Therefore, increasing the down sampling rate is not suitable for the detection of weld defects. In this paper, dilated convolution will be used to solve the problem of insufficient receptive field, and different dilated rates will be used to solve the problem of grid and abnormal aspect

ratio defects.

3 DRepDet Model and Algorithm

In this section, we first analyze the receptive field required by weld defects, then introduce RepPoints network architecture and the improved DRepDet architecture, and finally describe the DRepDet algorithm in detail.

3.1 Enlarge Receptive Field

In a typical neural network structure, the value of each output node of the fully connected (FC) layer depends on all inputs of the FC layer, while the value of each output node of the convolution layer depends on only one area of the convolution layer input. Other input values outside this area will not affect the output value, which is the receptive field. The pixels in the image area outside the receptive field area will not affect the feature vector on the feature map, so it is unlikely that the neural network only relying on a feature vector can find the object outside the corresponding input receptive field.

General tasks require that the larger the receptive field, the better. For example, the receptive field of the last convolution layer in image classification should be larger than that in the input image. The deeper the network depth, the larger the receptive field, and the better the performance. Dense prediction tasks require that the

Table 1 Comparison of different depth models of ResNet

Model depth	Down sampling rate/times	Receptive field/pixel	Parameter /10 ⁶	FLOPs/ Gb·s ⁻¹
ResNet50	32	483	25.5	3.8
ResNet101	32	1 027	44.5	7.6

receptive field of output feature map is large enough to ensure that important information is not ignored when making decisions. Generally, the deeper the better. The anchor boxes preset in the object detection task should strictly match the receptive field. If the anchor box is too large or deviates from the receptive field, it will seriously affect the detection performance.

The calculation formula of receptive field is as follows:

$$j_{\text{out}} = j_{\text{in}} * s \quad (1)$$

$$r_{\text{out}} = r_{\text{in}} + ((k + (k - 1) * (dr - 1)) - 1) * j_{\text{in}} \quad (2)$$

s is the step size of the convolution layer, j_{out} is the overall jump of the output characteristic graph, which is equal to the jump of the input characteristic graph, j_{in} times the step of the current convolution layer s , k is the size of the convolution kernel, dr is the dilated rate of the convolution kernel, r_{in} is the receptive field of the input characteristic graph, and r_{out} is the receptive field of the output characteristic graph.

We have made statistics on the defect area, width and height of the training set according to the size of the receptive field, as shown in Table 2. The receptive field is increased from 675 to 867. It can be seen from Table 2 that the receptive field of most defects is below 291×291 . The large-size defects are mostly harmful defects. In the field of industrial defect detection, the detection of hazardous defects is more important. The detection of this part of defects means that the products are unqualified. Therefore, the detection of this part of defects is particularly important. Ref. [21] proposed the effective

receptive field (ERF) theory. The paper found that not all pixels in the receptive field contribute the same to the output vector. In many cases, the distribution of pixels in the receptive field region follows the Gaussian distribution, the effective receptive field is only a part of the rational receptive field, and the Gaussian distribution decays rapidly from the center to the edge. In order to improve the detection performance of large-scale defects, it is necessary to output the feature map with a larger receptive field.

3.2 DRepDet Architecture

The overall architecture of DRepDet network is illustrated in Fig.3. It is modified on the basis of anchor-free detection network RepPoints.

The detection head (bounding box regression and classification) of DRepDet network is the same as RepPoints. It is an anchor-free detection framework. It represents the object as a set of points for positioning and recognition, while most object detectors rely on rectangular bounding box for detection and recognition. The bounding box indicates that only the rectangular spatial area of the target is considered, without considering the shape, posture and the location of the semantically important local area. In order to overcome the above disadvantages, RepPoints models a group of adaptive sampling points, which will automatically adjust to the bounding position of the object. In the training process, this set of sampling points is used to generate a pseudo box, which is compared with the ground truth box to calculate the loss.

In Section 3.1, we statistically analyze the receptive

Table 2 Receptive field statistics

Statistical scale	Statistics by area			Statistics according to the maximum value of width and height		
	Number of defects	Defect type	Receptive field ratio/%	Number of defects	Defect type	Receptive field ratio/%
<20	14 151	a, b	23.8	9 755	a, b	16.4
35	25 275	a, b, e, f, g	42.5	28 115	a, b, e, f, g	47.3
99*	8 069	a, b, c, d, e, f, g	13.6	5 715	a, b, c, d, e, f, g	9.6
291*	11 624	a, b, c, d, e, f, g	19.6	15 100	a, b, c, d, e, f, g	25.4
483*	201	a, b, c, d, e, f, g	0.3	257	a, b, c, d, e, f, g	0.4
547*	35	a, b, c, d, e	0.1	40	b, c, d, e, f, g	0.1
675*	25	b, c, d, e, f,	0.0	84	b, c, d, e, g	0.1
>675	34	b, c, d, e, f,	0.1	348	a, b, c, d, e, g	0.6

a: porosity; b: slag inclusion; c: intensive dispersed; d: intensive chain; e: lack of fusion; f: lack of penetration; g: crack; * is the receptive field of RepPoints backbone

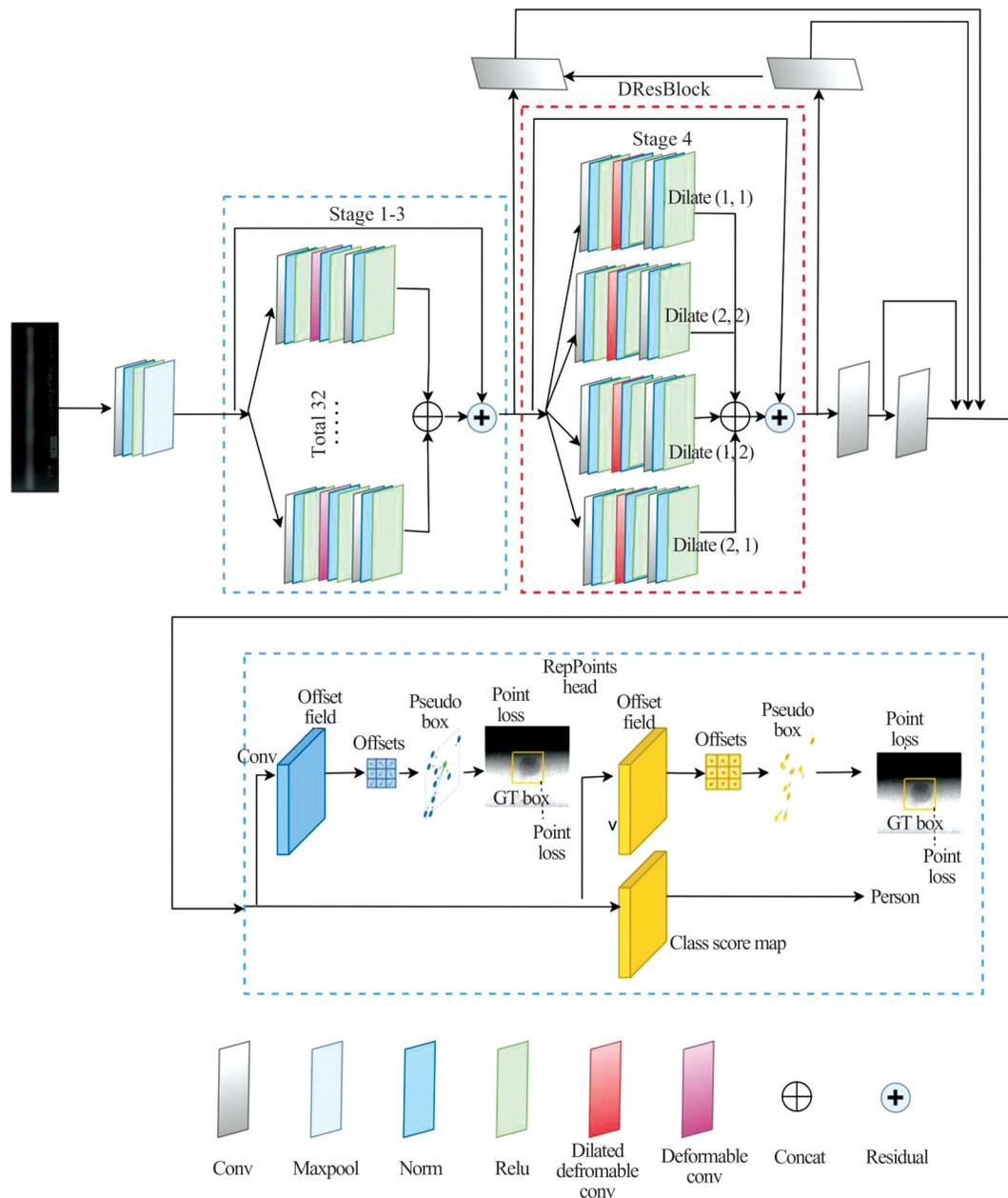


Fig. 3 DRRepDet architecture diagram

field of RepPoints and the scale of defects. The receptive field of the original backbone network of RepPoints cannot meet the distribution requirements of weld defects. Therefore, we design a multi branch convolution module DResBlock with different dilated rates to solve the problem that the scale of defects changes greatly, but the receptive field is not large enough. The backbone network of RepPoints is based on RetinaNet, generating five pyramid levels from stage 3 (down sampling rate of 8) to stage 7 (down sampling rate of 128). In order to expand the receptive field, we replace stage 4 with DResBlock. As shown in the dashed box at top center in Fig.3, this module uses four groups of convolutions with dilated rates (dr) of (1,1), (2,

2), (1,2), (2,1), and output after concatenating operation to integrate the feature of different scales.

3.3 DRRepDet Algorithm

The core of DRRepDet algorithm is the DResBlock module, which is similar to the residual module of ResNeXt network. Each branch of the residual module of ResNeXt network uses convolution kernels with the same size, width and height, and the extracted features cannot well represent objects with large differences in size and shape. The DResBlock module we proposed uses multiple groups of group convolution with different dilated rates to solve this problem.

When designing the DResBlock module, we adopted the bottleneck design similar to ResNeXt, mainly considering the following schemes (In these four schemes of DResBlock and corresponding figures, d denotes dimension): 1) Four groups of $32 \times 4d$ grouping convolutions with different dilated rates, and the results of the four groups of convolutions are added element-wise and added with the residual as the output, as shown in Fig.4 (a); 2) Four groups of $8 \times 4d$ grouping convolutions with different dilated rates. The input channel of each group is reduced by 1×1 convolution. The results of the four groups of convolutions are concatenated and the residual is added as the output, as shown in Fig.4 (b); 3) Four groups of $8 \times 4d$ group convolutions with different dilated rates. After 1×1 convolution, the channel dimension is directly divided into 4 groups. The results of the 4 groups of convolutions are concatenated and

add the residual as the output, as shown in Fig.4 (c); 4) Four groups of $32 \times 4d$ grouping convolution with different dilated rates. After concatenated, the 4 groups of results are output after 1×1 convolution and residual connect, as shown in Fig.4 (d).

The purpose of DResBlock module is to expand the receptive field to match the size of weld defects. The receptive fields obtained by the above four schemes are the same, but the model complexity and model performance are slightly different. We conduct a detailed experimental comparison in Section 4.2.

The backbone network of DRepDet algorithm is built based on DResBlock and follows the ResNeXt architecture. There is a stem module and four convolution stages. Each stage is composed of 3, 4, 6 and 3 bottleneck convolution blocks, respectively. From the second convolution stage, the convolution layer of 3×3

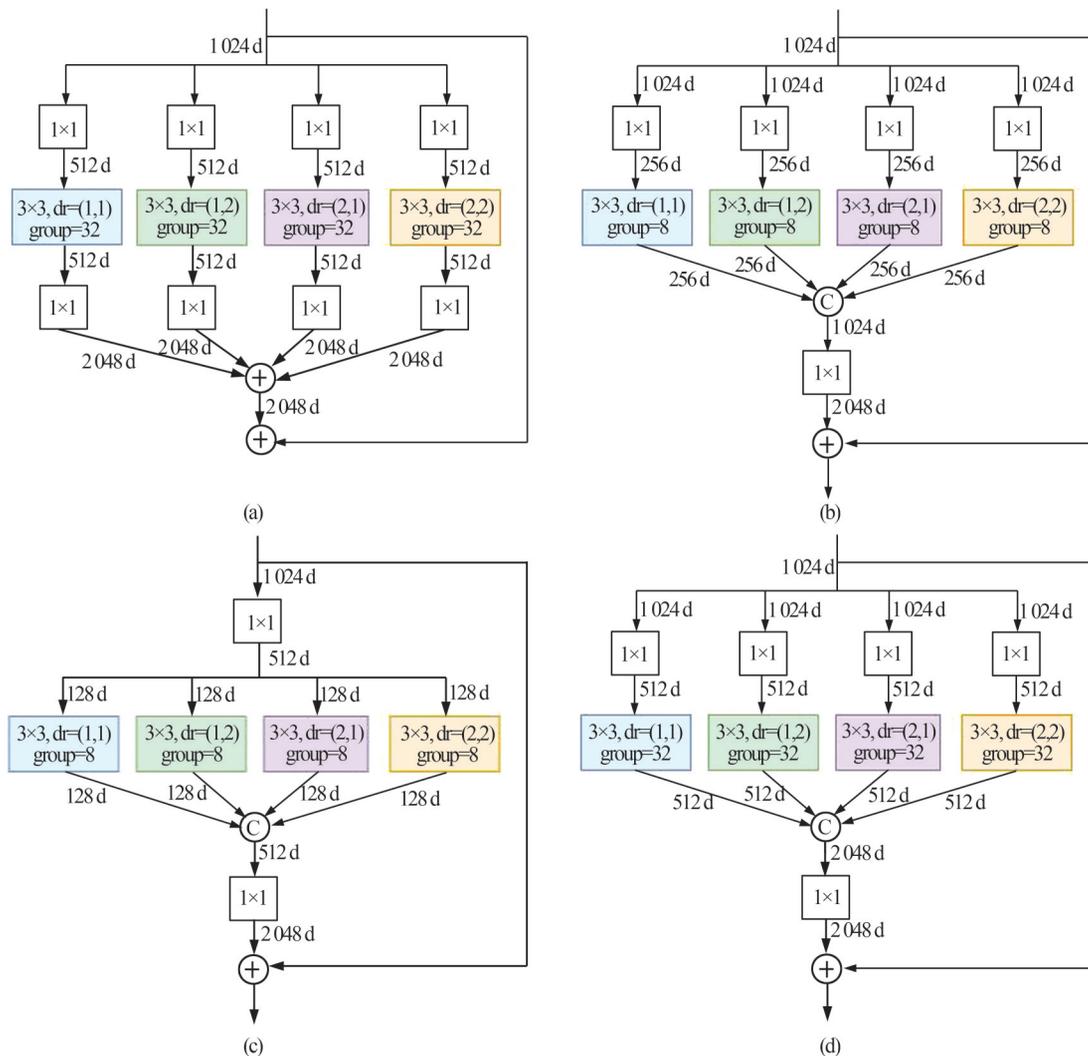


Fig. 4 DResBlock

uses deformable convolution, and the last stage uses DResBlock. The output of Feature Pyramid Networks (FPN) starts from the second convolution stage, and finally two additional FPN output layers are added. FPN has a total of 5 output layers. The feature maps of five different scales output by FPN are input to the regression and classification head and output the final classification and bounding box regression results.

3.4 Loss Function

DRepDet is a one-stage detector, and there is no Region Proposal Network (RPN). Positive and negative samples are extremely imbalanced. To alleviate this imbalance, the classification loss in training one stage detector mostly uses Focal Loss. We follow RetinaNet in supervising classification. SmoothL1Loss is a common regression loss for two-stage detectors, but this loss is sensitive to scale. When SmoothL1Loss is the same, IoU may vary greatly, which has a great impact on one-stage detectors without candidate proposal mechanism. Therefore, during the training of DRepDet, we choose GIoU loss to supervise the bounding box regression. GIoU loss is used both in the initial stage and the refine stage.

$$FL(p_i) = -\alpha_i (1 - p_i)^\gamma \log(p_i) \quad (3)$$

where α_i is a balance variant and γ is a tunable focusing parameter to adjust the rate at which easy examples are down-weighted, p_i is the predicted classification score.

$$\begin{cases} \text{IoU} = \frac{|A \cap B|}{|A \cup B|} \\ \text{GIoU} = \text{IoU} - \frac{|CA \cup B|}{|C|} \\ L_{\text{GIoU}} = 1 - \text{GIoU} \end{cases} \quad (4)$$

where A and B are bounding boxes, C is the smallest enclosing object of A and B , $CA \cup B$ is the area occupied by C excluding A and B .

The final loss L is a weighted sum of classification loss L_{class} and the two stage regression loss ($L_{\text{regress-init}}$, $L_{\text{regress-refine}}$), as in formula (5).

$$L = \lambda_1 L_{\text{class}} + \lambda_2 L_{\text{regress-init}} + \lambda_3 L_{\text{regress-refine}} \quad (5)$$

4 Experiment

4.1 Data Sets and Settings

We have conducted extensive experiments on our private weld defect dataset to validate the effectiveness of the algorithm proposed in this paper. The weld defect images are collected from the on-field pressure vessel,

with a resolution of about $8\,000 \times 2\,000$. The dataset includes 5 400 annotated images, 80% for training and 20% for validating. Due to the huge image size, $4\,000 \times 1\,000$ cutting is performed on the training set, and the overlap rate is 50%. Image cutting is not performed during inference.

We use the stochastic gradient descent (SGD) optimizer for training. The batch size is 8,4 GPUs (two pictures per GPU). The pretraining model on the COCO dataset is used for fine tuning, the initial value of the learning rate is 0.000 5, and the learning rate scheduling follows the "2×" setting^[31]. Data augmentation during training only uses random horizontal flip with probability of 0.5, and data augmentation is not used in inference process.

4.2 Experimental Result

4.2.1 Comparison of different dilated convolution experiments

In Table 3, we list the accuracy and recall of small defects and large defects under different dilated rates. The precision of oversized defects and defects with large height/width ratio is increased from 60.4% to 63.5%. The AP50 and Recall50 of big defects are improved by 3.1% and 3.3%, respectively. It can be seen from Table 3 that after using dilated convolution, the performance of small defects is hardly affected, while the performance of large defects is greatly improved. Since most of the large defects are hazardous and the number of these samples is small, the overall performance is not as good as the small defects. There are many long strip defects in hazardous defects, so we design the dilated rate with different width and height. Because there are both horizontal and vertical strip defects, and only one type of dilated convolution can improve the defect performance of the corresponding shape, we fuse the characteristics of different dilated rates. The fused features significantly improve the overall detection performance.

4.2.2 Comparison of four schemes of DResBlock

We select mAP50, mRecall50, FLOPs and params as the evaluation metrics of the four schemes of DResBlock. mAP50 is a commonly used metric to evaluate the comprehensive ability of all classes. It is an average of AP50 values on all classes. Similarly, mRecall₅₀ is an average of Recall50 values on all classes. In addition, we use FLOPs and parameters to evaluate the computational complexity of different models. Table 4 lists the comparison of evaluation metric of four schemes of DResBlock. It shows that scheme (d) has relatively high

Table 3 Comparison of different dilated rates %

Dilated rate	Minor defect (<483)		Big defect (>483)	
	AP50	Recall50	AP50	Recall50
(1,1)	85.5	91.7	60.4	65.6
(2,2)	85.6	91.5	61.7	66.2
(2,1)	85.2	91.4	61.0	65.8
(1,2)	85.4	91.5	61.1	66.0
(1,1)+(2,2)	85.6	92.3	63.5	68.9
+(2,1)+(1,2)				

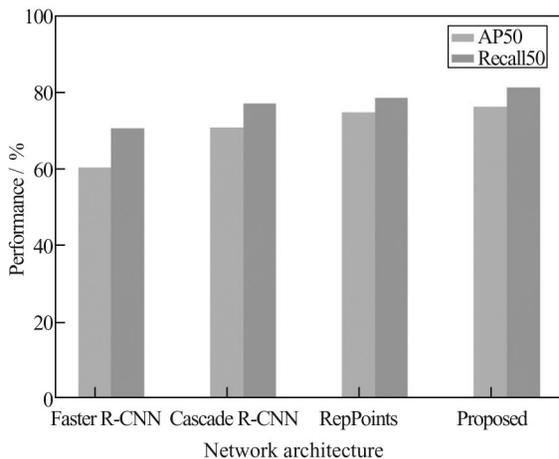
computation and memory complexity, but the accuracy and recall rate are better than other schemes. Since the industrial scene of weld defects requires higher accuracy, we choose scheme (d) for subsequent experiments if not specified.

Table 4 Comparison of four schemes of DResBlock

Scheme	mAP50/ %	mRecall50/ %	FLOPs/Gb·s ⁻¹	Params/10 ⁶
RepPoints	76.2	78.4	234.99	39.1
DResBlock(a)	77.1	79.1	281.11	77.93
DResBlock(b)	77.3	79.8	234.99	39.1
DResBlock(c)	77.2	79.6	234.99	39.1
DResBlock(d)	77.9	80.7	250.31	51.15

4.2.3 Comparison of various network architectures

The results of experiments on anchor-free networks, anchor-based networks and our proposed networks are illustrated in Fig. 5. Figure 5 shows that the AP50 and Recall50 of the network without anchor is higher than those

**Fig. 5** Comparison of different network architectures

of based on anchor. The reason is that the defect scale and aspect ratio of the weld defect vary greatly, and the preset anchor cannot cover such a large-scale range. The performance of our proposed network is further improved because the receptive field is adjusted according to the defects of large-scale and large aspect ratio of the dataset. The final performance of the whole network is improved a large margin, with 6% AP50 and 4.2% Recall50 compared with Cascade R-CNN and 1.4% AP50 and 2.9% Recall50 compared with RepPoints .

5 Conclusion

In this paper, we first carry out a multi-dimensional detailed analysis of weld defects, and find that weld defects had the problem of large difference in size and aspect ratio. To solve this problem, we analyze the receptive field of the network and study the method of expanding the receptive field. Based on these studies, a DResBlock module is proposed, which uses multi-branch convolution layers with different dilated rates to solve the problem of insufficient receptive fields. Based on DResBlock module, we design DRepDet model to detect weld defects. DRepDet can detect seven kinds of weld defects, and the receptive field is increased from 675 to 867. The precision of oversized defects and defects with large height/width ratio is increased from 60.4% to 63.5%, and the precision of defects with normal size is also improved.

References

- [1] National Boiler and Pressure Vessel Standardization Technical Committee (SAC/TC 262). NB/T 47013-2015. *Nondestructive Testing of Pressure Equipment* [S]. Beijing: National Energy Administration, 2015(Ch).
- [2] ISO/TC 44, Welding and Allied Processes, Subcommittee SC 10. EN ISO 5817:2014. *Welding - fusion Welded Joints of Steel, Nickel, Titanium and Their Alloys (except beam welding) - Quality Levels of Imperfections* [S]. Brussels: CEN-CENELEC Management Centre , 2014.
- [3] Hu J, Shen L, Albanie S, *et al*. Squeeze-and-excitation networks[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, **42**(8): 2011-2023.
- [4] He K M, Zhang X Y, Ren S Q, *et al*. Deep residual learning for image recognition [C]//2016 *IEEE Conference on Computer Vision and Pattern Recognition*. New York: IEEE, 2016: 770-778.

- [5] Xie S N, Girshick R, Dollár P, *et al.* Aggregated Residual Transformations for Deep Neural Networks [C]// *IEEE Conference on Computer Vision and Pattern Recognition*. New York: IEEE, 2017: 5987-5995.
- [6] Tan M X, Le Q V. EfficientNet: Rethinking model scaling for convolutional neural networks [C]// *Proceedings of the 36th International Conference on Machine Learning*. New York: IEEE, 2019, **97**: 6105-6114.
- [7] Redmon J, Divvala S, Girshick R, *et al.* You only look once: Unified, real-time object detection [C]//2016 *IEEE Conference on Computer Vision and Pattern Recognition*. New York: IEEE, 2016: 779-788.
- [8] Liu W, Anguelov D, Erhan D, *et al.* SSD: Single shot multi-box detector[C]// *Computer Vision – ECCV 2016*. Berlin: Springer-Verlag, 2016, **9905**: 21-37.
- [9] Ren S Q, He K M, Girshick R, *et al.* Faster R-CNN: Towards real-time object detection with region proposal networks[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, **39**(6): 1137-1149.
- [10] Cai Z W, Vasconcelos N. Cascade R-CNN: Delving into high quality object detection[C]//2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. New York: IEEE, 2018: 6154-6162.
- [11] He K M, Gkioxari G, Dollár P, *et al.* Mask R-CNN[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, **42**(2): 386-397.
- [12] Chen L C, Zhu Y K, Papandreou G, *et al.* Encoder-decoder with atrous separable convolution for semantic image segmentation[C]//*Computer Vision – ECCV 2018*. Berlin: Springer-Verlag, 2018, **11211**: 833-851.
- [13] Wang T, Chen Y, Qiao M N, *et al.* A fast and robust convolutional neural network-based defect detection model in product quality control[J]. *The International Journal of Advanced Manufacturing Technology*, 2018, **94**(9-12): 3465-3471.
- [14] Chen J W, Liu Z G, Wang H R, *et al.* Automatic defect detection of fasteners on the catenary support device using deep convolutional neural network[J]. *IEEE Transactions on Instrumentation and Measurement*, 2018, **67**(2): 257-269.
- [15] Tao X, Zhang D P, Ma W Z, *et al.* Automatic metallic surface defect detection and recognition with convolutional neural networks[J]. *Applied Sciences*, 2018, **8**(9): 1575-1589.
- [16] Lin T Y, Maire M, Belongie S, *et al.* Microsoft COCO: Common objects in context[C]//*Computer Vision – ECCV 2014*. New York: IEEE, 2014, **8693**: 740-755.
- [17] Lin T Y, Goyal P, Girshick R, *et al.* Focal loss for dense object detection[J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2020, **42**(2): 318-327.
- [18] Law H, Deng J. CornerNet: Detecting objects as paired keypoints[J]. *International Journal of Computer Vision*, 2020, **128**(3): 642-656.
- [19] Zhou X Y, Zhuo J C, Krähenbühl P. Bottom-up object detection by grouping extreme and center points[C]//2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. New York: IEEE, 2019: 850-859.
- [20] Yang Z, Liu S H, Hu H, *et al.* RepPoints: Point set representation for object detection[C]//2019 *IEEE/CVF International Conference on Computer Vision*. New York: IEEE, 2019: 9656-9665.
- [21] Luo W J, Li Y J, Urtasun R, *et al.* Understanding the effective receptive field in deep convolutional neural networks [C]// *Advances in Neural Information Processing Systems (NIPS)*. New York: IEEE, 2016: 4898-4906.
- [22] Chen L C, Papandreou G, Kokkinos I, *et al.* DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018, **40**(4): 834-848.
- [23] Chen L C, Papandreou G, Schroff F, *et al.* Rethinking atrous convolution for semantic image segmentation[EB/OL]. [2017-12-05]. <https://www.arXiv:1706.05587>.
- [24] Szegedy C, Vanhoucke V, Ioffe S, *et al.* Rethinking the inception architecture for computer vision[C]//2016 *IEEE Conference on Computer Vision and Pattern Recognition*. New York: IEEE, 2016: 2818-2826.
- [25] Liu S T, Huang D, Wang Y H. Receptive field block net for accurate and fast object detection[C]//*Computer Vision – ECCV 2018*. New York: IEEE, 2018, **11215**: 404-419.
- [26] Wu T Y, Tang S, Zhang R, *et al.* Tree-structured kronecker convolutional network for semantic segmentation[C]//2019 *IEEE International Conference on Multimedia and Expo*. New York: IEEE, 2019: 940-945.
- [27] Hu Y N, Wang J, Zhu Y Q, *et al.* Automatic defect detection from X-ray scans for aluminum conductor composite core wire based on classification neural network[J]. *NDT & E International*, 2021, **124**: 102549.
- [28] Ferguson M, Ak R, Lee Y T T, *et al.* Detection and segmentation of manufacturing defects with convolutional neural networks and transfer learning[J]. *Smart and Sustainable Manufacturing Systems*, 2018, **2**(1): 210-237.
- [29] Tokime R B, Maldague X, Perron L. Automatic defect detection for X-ray inspection: Identifying defects with deep convolutional network [C]// *Canadian Institute for Non-destructive Evaluation (CINDE)*. New York: IEEE, 2019.
- [30] Duan F, Yin S F, Song P P, *et al.* Automatic welding defect detection of X-ray images by using cascade Adaboost with penalty term[J]. *IEEE Access*, 2019, **7**: 125929-125938.
- [31] Detectron. Meta Research [DB/OL]. [2022-02-16]. <https://github.com/facebookresearch/Detectron>. □